

The Evolution of Multilocus Systems Under Weak Selection

Thomas Nagylaki

Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637

Manuscript received October 23, 1992

Accepted for publication February 19, 1993

ABSTRACT

The evolution of multilocus systems under weak selection is investigated. Generations are discrete and nonoverlapping; the monoecious population mates at random. The number of multiallelic loci, the linkage map, dominance, and epistasis are arbitrary. The genotypic fitnesses may depend on the gametic frequencies and time. The results hold for $s \ll c_{\min}$, where s and c_{\min} denote the selection intensity and the smallest two-locus recombination frequency, respectively. After an evolutionarily short time of $t_1 \sim (\ln s)/\ln(1 - c_{\min})$ generations, all the multilocus linkage disequilibria are of the order of s [i.e., $O(s)$ as $s \rightarrow 0$], and then the population evolves approximately as if it were in linkage equilibrium, the error in the gametic frequencies being $O(s)$. Suppose the explicit time dependence (if any) of the genotypic fitnesses is $O(s^2)$. Then after a time $t_2 \sim 2t_1$, the linkage disequilibria are nearly constant, their rate of change being $O(s^2)$. Furthermore, with an error of $O(s^2)$, each linkage disequilibrium is proportional to the corresponding epistatic deviation for the interaction of additive effects on fitness. If the genotypic fitnesses change no faster than at the rate $O(s^3)$, then the single-generation change in the mean fitness is $\Delta\bar{W} = \bar{W}^{-1}V_g + O(s^3)$, where V_g designates the genic (or additive genetic) variance in fitness. The mean of a character with genotypic values whose single-generation change does not exceed $O(s^2)$ evolves at the rate $\Delta\bar{Z} = \bar{W}^{-1}C_g + O(s^2)$, where C_g represents the genic covariance of the character and fitness (i.e., the covariance of the average effect on the character and the average excess for fitness of every allele that affects the character). Thus, after a short time t_2 , the *absolute* error in the fundamental and secondary theorems of natural selection is small, though the *relative* error may be large.

TWO natural approaches [both reviewed in EWENS (1979) Ch. 6 and 7, HASTINGS (1989), and NAGYLAKI (1992a) Ch. 8 and 10] have been fruitfully used to study selection at multiple loci. In most of this work, a panmictic, monoecious population with discrete, nonoverlapping generations and viability selection with constant genotypic viabilities are posited.

In the first approach, the great complexity and difficulty of the problem are reduced by imposing restrictions on the fitness pattern. Natural and intensely investigated restrictions have been that fitnesses be additive or multiplicative between loci or that they satisfy certain symmetries. Since these simplified problems are still quite complicated and difficult, most studies have focused on finding the equilibria, testing their local stability, and hence drawing conclusions about the local behavior of the linkage disequilibria and the mean fitness. Even if there are only two loci, stable linkage disequilibrium can be generated, and the mean fitness can decrease (MORAN 1964; KIMURA 1965; NAGYLAKI 1977a).

For additive fitnesses, there exist general, global results. EWENS (1969a,b) has extended to multiple loci the classic single-locus result that the mean fitness is

nondecreasing (MULHOLLAND and SMITH 1959; SCHEUER and MANDEL 1959; ATKINSON, WATTERSON and MORAN 1960; KINGMAN 1961). The gametic frequencies always converge to a stationary point in linkage equilibrium (KARLIN and FELDMAN 1970; KARLIN 1978; KUN and LYUBICH 1979). The fairly simple formula for the change in mean fitness can be written in the form

$$\Delta\bar{W} = \frac{V_g}{\bar{W}} (1 + E), \quad (1)$$

where V_g denotes the genic (or additive genetic) variance in fitness, and the relative error E is zero in the absence of dominance (NAGYLAKI 1989). There are several upper bounds on E , of which the most illuminating (though not the sharpest) is $|E| \leq \frac{1}{2}s$, where s designates the greatest multilocus selection coefficient (NAGYLAKI 1991). Thus, FISHER's (1930) fundamental theorem of natural selection holds approximately for weak selection:

$$\Delta\bar{W} \approx V_g/\bar{W}. \quad (2)$$

In the second approach, the fitness pattern is arbitrary. EWENS (1979, Ch. 8 and 9) and HASTINGS (1989) review numerical studies. To derive analytical results, it is customary and biologically reasonable to

This paper is dedicated to JAMES F. CROW, who has shared and encouraged my interest in the fundamental theorem of natural selection for twenty years.

posit that selection is much weaker than recombination. Even then we must be satisfied with conclusions that are usually less detailed and sometimes more qualitative than those for particular selection schemes. Much of the recent research has been stimulated by and oriented toward quantitative genetics [BARTON and TURELLI (1991) and references therein].

For two loci, we have considerable general understanding of evolution under weak selection (NAGYLAKI 1976, 1977a,b, Sec. 8.2; 1992a, Sec. 8.2). The results hold for $s \ll c$, where c signifies the recombination frequency. After an evolutionarily short time of $t_1 \sim (\ln s)/\ln(1 - c)$ generations, the linkage disequilibria are of the order of s [i.e., $O(s)$ as $s \rightarrow 0$], and then the population evolves approximately as if it were in linkage equilibrium, the error in the gametic frequencies being $O(s)$. Suppose the explicit time dependence (if any) of the genotypic fitnesses is $O(s^2)$. Then after a time $t_2 \sim 2t_1$, the linkage disequilibria are nearly constant, their rate of change being $O(s^2)$. If the genotypic fitnesses change no faster than at the rate $O(s^3)$, then

$$\Delta \bar{W} = \bar{W}^{-1} V_g + O(s^3), \quad t \geq t_2, \quad (3)$$

as $s \rightarrow 0$, where t ($= 0, 1, 2, \dots$) represents time in generations.

Equation 3 means that, after a short time t_2 , the *absolute* error in the fundamental theorem of natural selection is small. Although this suggests that (2) should hold generically, exceptions would be precluded only if the *relative* error were small, as described above for additive fitnesses. From (3) we see that the mean fitness can decrease only if the genic variance is much smaller than s^2 , which is the case if and only if the absolute value of the gene-frequency change per generation is much smaller than s . The latter can result from symmetry conditions or proximity to an equilibrium (MORAN 1964; KIMURA 1965; NAGYLAKI 1977a). Stable cycling in two-locus models (AKIN 1979, 1982, 1983; HASTINGS 1981) provides the most striking example of the occasional failure of the fundamental theorem of natural selection.

AKIN (1979) and SHAHSHAHANI (1979) have developed an elegant mathematical formalism for a multi-locus continuous-time model in which gametic Hardy-Weinberg proportions are assumed, and PASSEKOV (1984) briefly discusses the weak-selection dynamics of this model. Although we expect this model to have usually the same qualitative behavior as the standard discrete model investigated in this paper, it is not easy to derive this continuous model rigorously in such a manner that it covers the parameter space of biological interest. Even for a single locus, the continuous Hardy-Weinberg model applies only to weak selection (NAGYLAKI and CROW 1974; NAGYLAKI 1976, 1977b, Sec. 4.10; 1992a, Sec. 4.10 and Problem 4.22). The continuous multilocus model can be deduced as the

limit of the discrete one if selection is weak and linkage is tight. However, if not all the recombination frequencies are small, then one must first prove one of the major results of this paper, *viz.*, that all the multilocus linkage disequilibria rapidly become $O(s)$ in the discrete model. For two loci, this argument is outlined in Problem 8.15 in NAGYLAKI (1992a).

In principle, a continuous model that properly incorporates deviations from Hardy-Weinberg proportions is exact for a continuously reproducing population if either it has no age dependence or it has reached a stable age distribution. Such a model has been formulated and analyzed for two loci (NAGYLAKI and CROW 1974; NAGYLAKI 1976, 1977b, Sec. 8.4; 1992a, Sec. 8.4); MOODY (1978) has formulated the general multilocus model.

The ideas and methods used in the weak-selection analysis of the two-locus system have been adapted and extended to several other biological situations: single autosomal and X-linked loci in a dioecious population (NAGYLAKI 1979a, 1992a, Sec. 7.4), density-dependent selection (NAGYLAKI 1979b, 1992a, Sec. 4.11), and selection on both viability and fertility (NAGYLAKI 1987). This work is briefly reviewed in NAGYLAKI (1991).

In this paper, we shall generalize the analysis of the two-locus system described above to an arbitrary number of loci, and, motivated by quantitative genetics, we shall extend (3) to an arbitrary character, thereby establishing the secondary theorem of natural selection. Previous investigations of this theorem were based on more restrictive assumptions: linear regression (ROBERTSON 1966, 1968; NAGYLAKI 1992b), normality (LANDE 1976), two loci in continuous time (CROW and NAGYLAKI 1976; NAGYLAKI 1989), no epistasis (NAGYLAKI 1989, 1991), purely additive genetics (TURELLI and BARTON 1990), or linkage equilibrium (NAGYLAKI 1992b).

We adhere to the canonical interpretation of the fundamental theorem of natural selection: $\Delta \bar{W}$ in (3) is the actual (or total) change in the mean fitness. PRICE (1972) and EWENS (1989) have argued cogently that FISHER (1930) really treated only a certain part of $\Delta \bar{W}$, for which (2) is exact under very mild assumptions. This interpretation is described and discussed in NAGYLAKI (1991). It might be most accurate to refer to their result as the *Fisher-Price-Ewens theorem on natural selection* and to (3) as the *asymptotic fundamental theorem of natural selection*.

In the next section, we shall formulate our problem and derive some preliminary results. In the following section, we shall prove that the multilocus linkage disequilibria rapidly become small. Then we shall demonstrate that the population evolves approximately as if it were in linkage equilibrium. In the succeeding section, we shall show that the linkage disequilibria rapidly become nearly constant. Next,

we shall decompose the gametic covariance of an arbitrary character and fitness. In the following section, we shall relate each linkage disequilibrium to the corresponding epistatic deviation for the interaction of additive effects on fitness. Then we shall prove the asymptotic fundamental and secondary theorems of natural selection. In the final section, we shall summarize and discuss our results.

In each section, we shall state and comment on our major results at the beginning and prove them in the following subsections.

FORMULATION

Here, we formulate our problem and derive some preliminary results.

In Table 1, we list and briefly define the symbols used in this paper. We refer to more precise definitions in the text and displays by equation number: thus, (7), (7)+, and (7)− mean that the definition appears in, below, and above Equation 7, respectively.

Generations are discrete and nonoverlapping; the monoecious population mates at random. The number of multiallelic loci, the linkage map, dominance, and epistasis are arbitrary. If there are fertility differences, the fertility of each mating can be expressed as the product of factors corresponding to the genotypes involved (BODMER 1965; NAGYLAKI 1977b, Sec. 8.1; 1992a, Sec. 8.1). The genotypic fitnesses may depend on the gametic frequencies and time.

Suppose there are n loci and M_k alleles $A_k^{(i)}$ ($i_k = 1, 2, \dots, M_k$) at locus k . We put $i = (i_1, i_2, \dots, i_n)$ and denote the frequency of the gamete $A_{i_1}^{(1)} A_{i_2}^{(2)} \dots A_{i_n}^{(n)}$ (which, for brevity, we shall call gamete i) immediately after gametogenesis by p_i ; collectively, these form the vector \mathbf{p} . Then the frequency of $A_k^{(i)}$ in gametes is

$$p_{i_k}^{(k)} = \sum_i^{(k)} p_i, \quad (4)$$

where the sum is over all components of the vector i except i_k .

Let $W_{ij}(\mathbf{p}, t)$ designate the fitness of genotype ij . We assume that $W_{ij}(\mathbf{p}, t)$ is bounded away from 0 and ∞ and is continuously differentiable with respect to \mathbf{p} for every i and j . Since this paper concerns weak selection, exclusion of lethality and sterility is not an additional restriction. We define the mean fitnesses of the gamete i , the allele $A_k^{(i)}$, and the population by

$$W_i(\mathbf{p}, t) = \sum_j W_{ij}(\mathbf{p}, t) p_j, \quad (5a)$$

$$p_{i_k}^{(k)} W_{i_k}^{(k)}(\mathbf{p}, t) = \sum_i^{(k)} W_i(\mathbf{p}, t) p_i, \quad (5b)$$

$$\bar{W}(\mathbf{p}, t) = \sum_{i,j} W_{ij}(\mathbf{p}, t) p_i p_j. \quad (5c)$$

We shall usually suppress the arguments \mathbf{p} and t . For both fitness and the arbitrary character Z , we shall consistently use lowercase letters to represent the de-

viations of quantities from their means:

$$w_{ij} = W_{ij} - \bar{W}, \quad (6a)$$

$$w_i = W_i - \bar{W} = \sum_j w_{ij} p_j, \quad (6b)$$

$$p_{i_k}^{(k)} w_{i_k}^{(k)} = p_{i_k}^{(k)} (W_{i_k}^{(k)} - \bar{W}) = \sum_i^{(k)} w_i p_i. \quad (6c)$$

Thus, w_i and $w_{i_k}^{(k)}$ signify the average excesses for fitness of the gamete i and the allele $A_k^{(i)}$, respectively.

To specify the effect of recombination, we discretize the formulation of FLEMING (1979). Let $N = \{1, 2, \dots, n\}$ denote the set of loci. Throughout this paper, I designates a proper subset of N including one (*i.e.*, $1 \in I \subset N$) and $J = N - I$ is its complement, whereas R designates a nonempty subset of N (*i.e.*, $R \subseteq N$, unless R represents a proper subset, in which case $R \subset N$) and $Q = N - R$. We write i_R for the vector with components i_k for every k in R . Let c_I signify the probability of reassociation of the genes at the loci in I , inherited from one parent, with the genes at the loci in J , inherited from the other.

Prominent roles will be played by the total recombination frequency.

$$c_{\text{tot}} = \sum_I c_I \quad (7)$$

and the recombination frequency c_{kl} between loci k and l such that $k < l$. To calculate the two-locus recombination frequencies c_{kl} in terms of the linkage map c_I , define the set of sets

$$N_{kl} = \{I: k \in I \text{ and } l \in J, \text{ or } k \in J \text{ and } l \in I\}. \quad (8)$$

Then we have

$$c_{kl} = \sum_{I \in N_{kl}} c_I. \quad (9)$$

The most important recombination parameter is the smallest two-locus frequency

$$c_{\min} = \min c_{kl}, \quad (10)$$

in which the minimum is over every $k, l \in N$ such that $k < l$.

The gametic frequencies satisfy the recursion relations

$$\Delta p_i = \bar{W}^{-1} p_i w_i - D_i, \quad (11a)$$

where

$$D_i = \bar{W}^{-1} \sum_j \sum_I c_I (W_{ij} p_i p_j - W_{i_j j_i} p_{i_j} p_{j_i}) \quad (11b)$$

denotes a measure of linkage disequilibrium in gamete i , and $p_{i_j j_i}$ designates the frequency of the gamete formed from the genes i_j , inherited from one parent, and the genes j_i , inherited from the other.

To derive the recursion relation for the gene frequencies $p_{i_k}^{(k)}$, we must prove the identity

$$\sum_i^{(k)} D_i = 0 \quad (12)$$

TABLE I
Glossary of symbols

Symbol	Reference	Definition
$A_{ik}^{(k)}$	(4)–	Allele i_k at locus k
a_i	(47b)	Constant
a_R	(140b)	Constant
B_i	(48)+	Constant
b_i	(47b)	Constant
$b_{S,M}$	(155b)	Constant
C_{gam}	(105a)	Gametic covariance of Z and W
C_g	(105b)	Genic covariance of Z and W
C_e	(105c)	Epistatic covariance of Z and W
$C_e^{(R)}$	(109)	Component of epistatic covariance of Z and W due to the interaction of all the loci in R
c	(3)–	Two-locus recombination frequency
c_I	(7)–	Frequency of reassociation of the genes at the loci in I , inherited from one parent, with the genes at the loci in J , inherited from the other
c_{tot}	(7)	Total frequency of recombination
c_{kl}	(7)+	Recombination frequency between loci k and l such that $k < l$
c_{min}	(10)	The smallest two-locus recombination frequency
$c_K^{(R)}$	(133)	Frequency of reassociation of the genes at the loci in K , inherited from one parent, with the genes at the loci in $R - K$, inherited from the other
$c_{tot}^{(R)}$	(134)	Total frequency of recombination among the loci in R
D_i	(11b)	n -locus linkage disequilibrium: the difference between the recombined and unrecombined adult genotypic frequencies, weighted by the recombination frequencies and summed over recombination events and one of the gametes
$D_{iR}^{(R)}$	(123)	Linkage disequilibrium among all the loci in R : defined for R as is D_i for N
$d_{iR}^{(R)}$	(17a)	Linkage disequilibrium among all the loci in R : defined for R as is d_i for N
d_i	(18a)	n -locus linkage disequilibrium: the difference between the frequency of gamete i and the product of the corresponding allelic frequencies
$\tilde{d}_{iR}^{(R)}$	(71)	Scaled linkage disequilibrium among all the loci in R
\tilde{d}_i	(83)	Scaled n -locus linkage disequilibrium
E	(1)	Relative error in the fundamental theorem of natural selection
\mathcal{E}	(128)–	Expectation over the gametic frequencies p_i
e_i	(87a)	Total epistatic deviation in z_i
$e_{iR}^{(R)}$	(101)	Epistatic deviation in z_i due to interactions in all subsets of R
F_i	(151)	Recombination function
f_i	(34a)	Selection function
$f_{iR}^{(R)}$	(34d)+	Selection function for the loci in R
$G_{iR}^{(R)}$	(160b)	Epistatic function
g_i	(34d)	Recombination function
$g_{iR}^{(R)}$	(34d)+	Recombination function for the loci in R
$H_{i_k}^{(k)}$	(55)	Allelic selection function
$h_{i_k}^{(k)}$	(56)	Allelic selection function
I	(6c)+	Proper subset of loci $\{1, 2, \dots, n\}$ including 1
i_k	(4)–	Allelic index at locus k
i	(4)–	Gametic index (i_1, i_2, \dots, i_n)
i_R	(6c)+	Vector with components i_k for every k in R
J	(6c)+	$N - I$, the complement of I
j	(4)+	Gametic index
K	(53a)	Constant
K	(131)	Subset of R
k	(4)–	Locus index
$L_{iR}^{(k)}$	(61)–	Constant
L	(63)	Constant
L	(139)+	Subset of N
l	(7)+	Locus index
l	(139)+	$ L $, the number of elements of L
M_k	(4)–	The number of alleles at locus k
M	(57)+	Constant
M	(131)	Subset of N
m	(155b)+	$ M $, the number of elements of M
N	(6c)+	$\{1, 2, \dots, n\}$, the set of loci
N_{kl}	(8)	Set of subsets of I of N such that each I contains either k or l

Table I—Continued

Symbol	Reference	Definition
$N_K^{(R)}$	(131)	Set of proper subsets M of N such that each M contains K and $N - M$ contains $R - K$
n	(4)–	The number of loci
O	(3)–	Order symbol
p_i	(4)–	Frequency of gamete i
\mathbf{p}	(4)–	Vector of gametic frequencies
$p_i^{(k)}$	(4)	Frequency of the allele $A_i^{(k)}$
p_{ijj}	(11b)+	Frequency of the gamete formed from the genes i , inherited from one parent, and the genes j , inherited from the other
$p_{iR}^{(R)}$	(16)	Gametic frequency for loci in R
p	(24)+	Allelic frequency at a diallelic locus
Q	(6c)+	$N - R$, the complement of R
$q_{iR}^{(R)}$	(17b)	Product of the allelic frequencies at the loci in R
q_i	(18b)	Product of all the allelic frequencies
R	(6c)+	Subset of N
$r(\mathbf{p}, t)$	(20a)	Selection intensity in generation t with gametic frequencies \mathbf{p}
r	(140b)+	$ R $, the number of elements of R
r_1	(133)+	The smallest element of R
$r_k(\mathbf{p}, t)$	(180a)	Selection intensity at locus k in generation t with gametic frequencies \mathbf{p}
S	(99)	Subset of N
s	(20b)	Selection intensity
s^*	(180b)	The greatest single-locus selection intensity
T	(147)+	$N - S$, the complement of S
t	(3)+	Time in generations
t_1	(39)	Characteristic time for reduction of the linkage disequilibria to $O(s)$
t_2	(70)	Characteristic time to approximate constancy of the linkage disequilibria
t_R	(69)+	Characteristic time for reduction of the linkage disequilibria for loci in R
t_N	(69)+	t_1
t_3	(70)+	Characteristic time for convergence to an equilibrium point
u_i	(104a)	Total epistatic deviation in w_i
$u_{iR}^{(R)}$	(124)	Epistatic deviation in w_i due to interactions in all subsets of R
V_{gam}	(88a)	Gametic variance in Z
V_g	(88b)	Genic variance in Z
V_e	(88c)	Epistatic variance in Z
V	(95)	Total genetic variance in Z
$V_{\text{gam,dom}}$	(95)	Variance due to nonadditivity of the gametic effects on Z
$V_e^{(R)}$	(97)	Epistatic variance in Z due to the interaction of all the loci in R
V_{A^*}	(117)–	Epistatic variance due to all sets of r additive effects on Z
$v_{iR}^{(R)}$	(108)+	Epistatic deviation in w_i due to the interaction of all the loci in R
W_{ij}	(4)+	Fitness of genotype ij
W_i	(5a)	Fitness of gamete i
$W_i^{(k)}$	(5b)	Fitness of allele $A_i^{(k)}$
\bar{W}	(5c)	Mean fitness
W	(84)+	Fitness
w_{ij}	(6a)	$W_{ij} - \bar{W}$
w_i	(6b)	$W_i - \bar{W}$
$w_i^{(k)}$	(6c)	$W_i^{(k)} - \bar{W}$
$X_i^{(k)}$	(13b)	Sum in (13b)
$x(t)$	(59)	Total absolute deviation of the allelic frequencies
$x_{iR}^{(R)}$	(96)+	Epistatic deviation in z_i due to the interaction of all the loci in R
$y_{ij}^{(j)}$	(143b)	Factor in (143a)
$y(t)$	(64)	Upper bound on $x(t)$
Z	(5c)+	Trait value
Z_{ij}	(85a)–	Trait value of genotype ij
Z_i	(85a)	Trait value of gamete i
$Z_i^{(k)}$	(85b)	Trait value of allele $A_i^{(k)}$
\bar{Z}	(85c)	Mean trait value
z_{ij}	(86a)	$Z_{ij} - \bar{Z}$
z_i	(86b)	$Z_i - \bar{Z}$
$z_i^{(k)}$	(86c)	$Z_i^{(k)} - \bar{Z}$
$\alpha_{iR}^{(R)}$	(37)	Constant
α_i	(39)–, (51b)	$\alpha_i^{(N)}$

Table I—Continued

Symbol	Reference	Definition
$\alpha^{(N)}$	(39)–	$\max_i \alpha_i$
$\beta_{iR}^{(R)}$	(37)	Constant
β_i	(51b)	$\beta_i^{(N)}$
Γ_i	(82)+	Constant
γ_N	(34b)	$1 - c_{\text{tot}}$, the probability that there is no recombination
γ_R	(34d)+	$1 - c_{\text{tot}}^{(R)}$, the probability that there is no recombination among the loci in R
Δ	(1)	Operator for change in one generation
$\Delta_{iR}^{(R)}$	(128)	Product-moment linkage disequilibrium among all the loci in R
Δ_i	(150)	Product-moment n -locus linkage disequilibrium
δ_i	(49)	Upper bound on $ d_i $
ζ_i	(87a)	Total additive effect on z_i
$\zeta_{ik}^{(k)}$	(87b)	Average effect of $A_{ik}^{(k)}$ on Z
$\tilde{\zeta}_{ik}^{(k)}$	(91)+	Translated value of $\zeta_{ik}^{(k)}$
$\tilde{\zeta}_i$	(91)+	Translated value of ζ_i
η_i	(81)	Constant
θ_i	(80)	Constant
$\kappa_{iR}^{(R)}$	(69)	Constant
κ_i	(70)–	$\kappa_i^{(N)}$
$\kappa^{(N)}$	(70)–	$\max_i \kappa_i$
κ	(153)+	$ K $, the number of elements in K
Λ_i	(78)	Constant
λ_R	(37)+	The complement of the smallest two-locus recombination frequency in R
λ_N	(38)	$1 - c_{\text{min}}$, the complement of the smallest two-locus recombination frequency
$\mu_{iR}^{(R)}$	(69)	Constant
μ_i	(84)	$\mu_i^{(N)}$
ν_i	(81)	Constant
ξ_i	(104a)	Total additive effect on w_i
$\xi_{ik}^{(k)}$	(104b)	Average effect of $A_{ik}^{(k)}$ on W
$\pi_{ik}^{(k)}$	(52)–	Frequency of the allele $A_{ik}^{(k)}$ on the linkage-equilibrium surface
π_i	(53b)	Frequency of gamete i on the linkage-equilibrium surface
$\boldsymbol{\pi}$	(53b)+	Vector of gametic frequencies on the linkage-equilibrium surface
ρ_i	(34c)	Constant
$\rho_{iR}^{(R)}$	(35b)	Constant
$\Sigma_i^{(k)}$	(4)	Sum over all components of the vector i except i_k
σ	(25)	Fitness parameter
σ	(153)+	$ S $, the number of elements of S
σ_k	(91)+	Constant
τ	(42)	Summation index for time
τ	(150)+	$ T $, the number of elements of T
φ_S	(154b)	Recombination function
$\chi_{ik}^{(k)}$	(127)	Indicator function for the allele $A_{ik}^{(k)}$
$\chi_{ik}^{(K)}$	(138)	Indicator function for gamete i_k in K
ψ_S	(154c)	Recombination function
$\omega_{R,S}$	(161b)	Sum in (161b)
prime	(34a)	Generation $t + 1$
sup	(20b)	Supremum, or least upper bound
\approx	(2)	Approximate equality
\sim	(3)–	Asymptotic equality
\in	(6c)+	An element of
\subset	(6c)+	A proper subset of
\subseteq	(6c)+	A subset of
$ L $	(139)+	For a set L , this signifies the number of elements in L

For both the Roman and Greek alphabets, uppercase letters precede lowercase ones. For each uppercase or lowercase letter, listing is in order of appearance of the precise definition in the text. The references are to the equation closest to the precise definition of each symbol. Thus, (7), (7)+, and (7)– would refer to Equation 7, the text below Equation 7, and the text above Equation 7, respectively.

for every k , i.e., the single-locus "linkage disequilibria" are zero. From (11b), (5a), and (5b) we obtain

$$\bar{W} \sum_i^{(k)} D_i = \sum_i c_i [W_{i_k}^{(k)} p_{i_k}^{(k)} - X_{i_k}^{(k)}(I)], \quad (13a)$$

where

$$X_{i_k}^{(k)}(I) = \sum_i^{(k)} \sum_j W_{ijj,ji} p_{ijj,ji} p_{ji}^{(k)}. \quad (13b)$$

There are two possibilities in (13b): either $k \in I$ or $k \in J$. If $k \in I$, then (5a) and (5b) yield

$$\begin{aligned} X_{i_k}^{(k)}(I) &= \sum_{i_j}^{(k)} \sum_{j_l} \sum_{j_l} W_{ijj,ji} p_{ijj,ji} p_{ji}^{(k)} \\ &= W_{i_k}^{(k)} p_{i_k}^{(k)}; \end{aligned} \quad (14)$$

a similar argument produces the same result for $k \in J$. Therefore, (13a) establishes (12). Summing (11a) and appealing to (4), (6c), and (12), we find

$$\bar{W} \Delta p_{i_k}^{(k)} = p_{i_k}^{(k)} w_{i_k}^{(k)}. \quad (15)$$

Although it is the linkage disequilibria D_i that enter the basic recursion relations (11), a simpler set of linkage disequilibria that do not involve the fitnesses (and which seem not to have been used for more than two loci) will be more convenient for our analysis. As above, suppose $R \subseteq N$ and $Q = N - R$. Then the gametic frequencies for loci in R read

$$p_{i_R}^{(R)} = \sum_{i_Q} p_i; \quad (16)$$

of course, $p_{i_N}^{(N)} = p_i$. We introduce the linkage disequilibria

$$d_{i_R}^{(R)} = p_{i_R}^{(R)} - q_{i_R}^{(R)}, \quad (17a)$$

in which

$$q_{i_R}^{(R)} = \prod_{k \in R} p_{i_k}^{(k)}. \quad (17b)$$

In particular, for $R = N$ we have

$$d_i = p_i - q_i, \quad (18a)$$

in which

$$q_i = \prod_{k=1}^n p_{i_k}^{(k)}. \quad (18b)$$

The analog of (12) follows immediately from (18) and (4): for every k ,

$$\sum_i^{(k)} d_i = 0. \quad (19)$$

From (11b) and (18) we see that if there is no position effect and $d_i = 0$ for every i , then $D_i = 0$ for every i .

We now turn to the definition of the selection intensity s . At least for weak selection, a simple, natural definition is often manifest; for our purposes, definitions of the same order of magnitude are equivalent. To be specific, however, we choose the most

conservative definition. We take

$$r(\mathbf{p}, t) = [\max_{i,j} W_{ij}(\mathbf{p}, t) - \min_{i,j} W_{ij}(\mathbf{p}, t)] / \min_{i,j} W_{ij}(\mathbf{p}, t) \quad (20a)$$

as the selection intensity in generation t with gametic frequencies \mathbf{p} and define

$$s = \sup_{\mathbf{p}, t} r(\mathbf{p}, t). \quad (20b)$$

Observe that (6a), (5c), and (20) give

$$\frac{|w_{ij}(\mathbf{p}, t)|}{\bar{W}(\mathbf{p}, t)} \leq r(\mathbf{p}, t) \leq s \quad (21a)$$

for every i and j , which is equivalent to

$$\left| \frac{W_{ij}(\mathbf{p}, t)}{\bar{W}(\mathbf{p}, t)} - 1 \right| \leq r(\mathbf{p}, t) \leq s. \quad (21b)$$

Thus,

$$w_{ij}(\mathbf{p}, t) = O(s) \quad (22)$$

as $s \rightarrow 0$. Here and below, unless indicated otherwise, all order symbols are uniform in \mathbf{p} and t . If the fitnesses are independent of time, the supremum in (20b) becomes a maximum over the gametic frequencies. For constant fitnesses, (20) simplifies to

$$s = (\max_{i,j} W_{ij} - \min_{i,j} W_{ij}) / \min_{i,j} W_{ij}. \quad (23)$$

An important immediate consequence of (15) and (22) is that gene frequencies change slowly:

$$\Delta p_{i_k}^{(k)} = O(s). \quad (24)$$

Weak selection will mean $s \ll c_{\min}$.

Two examples may illuminate the definition of s . Consider a single diallelic locus with gene frequency p .

For the fitness pattern

$$W_{11} = 1 + 2\sigma^2, \quad W_{12} = 1 + \sigma, \quad W_{22} = 1, \quad (25)$$

with $\sigma > -1$, a natural choice for $|\sigma| \ll 1$ would be $s = |\sigma|$. This agrees with (23), from which we easily deduce

$$s = \begin{cases} (2\sigma^2 - \sigma)/(1 + \sigma), & \sigma < 0, \\ \sigma, & 0 \leq \sigma \leq 1/2, \\ 2\sigma^2, & \sigma > 1/2. \end{cases} \quad (26)$$

The large value of s in (26) for σ close to -1 correctly reflects the fact that the heterozygote is then almost lethal.

For the frequency-dependent fitness pattern

$$W_{11}(p) = 1 + p + \sigma(4 - 6p), \quad (27a)$$

$$W_{12}(p) = 1 + p + \sigma(2 - 4p), \quad (27b)$$

$$W_{22}(p) = 1 + p + \sigma(1 - p), \quad (27c)$$

with $0 < \sigma < 1$, a natural choice for $\sigma \ll 1$ would be

$s = \sigma$. Straightforward application of (20) leads to

$$s = \begin{cases} 3\sigma/(1 + \sigma), & 0 < \sigma \leq 1/2, \\ \sigma/(1 - \sigma), & 1/2 < \sigma < 1, \end{cases} \quad (28)$$

which is indeed of order σ for $\sigma \ll 1$. Note that $s \rightarrow \infty$ as $\sigma \rightarrow 1$ because

$$W_{11}(1) = W_{12}(1) = 2(1 - \sigma) \rightarrow 0 \quad (29)$$

as $\sigma \rightarrow 1$.

Our last task in this section is to derive a recursion relation for the linkage disequilibria d_i . From (18), (24), (11a), and (22) we obtain

$$\Delta d_i = \Delta p_i + O(s) \quad (30)$$

$$= -D_i + O(s). \quad (31)$$

Successively invoking (11b), (22), (16), and (17), we deduce

$$D_i = \sum_I c_I (p_i - p_{i_I}^{(I)} p_{i_I}^{(I)}) + O(s) \quad (32)$$

$$= \sum_I c_I (d_i - d_{i_I}^{(I)} d_{i_I}^{(I)} - d_{i_I}^{(I)} q_{i_I}^{(I)} - d_{i_I}^{(I)} q_{i_I}^{(I)}) + O(s). \quad (33)$$

Substituting (33) into (31) and recalling (7) produces the recursion relation

$$d'_i = \gamma_N d_i + s f_i(\mathbf{p}, t) + g_i(p_i), \quad (34a)$$

where the prime signifies generation $t + 1$;

$$\gamma_N = 1 - c_{\text{tot}} \quad (34b)$$

represents the probability that there is no recombination among the loci in N ; the function $f_i(\mathbf{p}, t)$ satisfies

$$|f_i(\mathbf{p}, t)| \leq \rho_i \quad (34c)$$

for some constant ρ_i independent of \mathbf{p} , t , and s , and $f_i(\mathbf{p}, t)$ is independent of t if the genotypic fitnesses are independent of t ; and

$$g_i(p_i) = \sum_I c_I (d_{i_I}^{(I)} d_{i_I}^{(I)} + d_{i_I}^{(I)} q_{i_I}^{(I)} + d_{i_I}^{(I)} q_{i_I}^{(I)}). \quad (34d)$$

Note that $0 < \gamma_N < 1$. Our analysis will be based on (15) and (34), rather than on (11).

The recursion relation for $d_{i_R}^{(R)}$, where $R \subseteq N$, can be written down directly because it must have the form (34), as can be confirmed by observing from (16), (17), and (18) that

$$d_{i_R}^{(R)} = \sum_{i_Q} d_i, \quad q_{i_R}^{(R)} = \sum_{i_Q} q_i$$

and summing (34a) over i_Q . Thus, we have

$$d_{i_R}^{(R)'} = \gamma_R d_{i_R}^{(R)} + s f_{i_R}^{(R)}(\mathbf{p}, t) + g_{i_R}^{(R)}(p_{i_R}^{(R)}), \quad (*)$$

in which γ_R denotes the probability that there is no recombination in R ; on account of the embedding $R \subseteq N$, the selection function

$$f_{i_R}^{(R)}(\mathbf{p}, t) = \sum_{i_Q} f_i(\mathbf{p}, t)$$

still depends on the full vector \mathbf{p} of n -locus gametic frequencies; and for suitably defined recombination frequencies in R , the recombination term $g_{i_R}^{(R)}$ has the form (34d).

We shall need (*), however, only for two-locus subsystems. In this case, since $d_k^{(k)} = 0$ for every k and i_k by (17), therefore (34d) tells us that $g_{i_R}^{(R)} = 0$, and hence (*) reduces to

$$d_{i_R}^{(R)'} = \gamma_R d_{i_R}^{(R)} + s f_{i_R}^{(R)}(\mathbf{p}, t), \quad (35a)$$

where

$$|f_{i_R}^{(R)}(\mathbf{p}, t)| \leq \rho_{i_R}^{(R)} \quad (35b)$$

for some constant $\rho_{i_R}^{(R)}$ independent of \mathbf{p} , t , and s . Now (32) gives

$$D_{i_R}^{(R)} = \sum_{i_Q} D_i = c^{(R)} d_{i_R}^{(R)} + O(s), \quad (36)$$

where $c^{(R)} = 1 - \gamma_R$ designates the recombination frequency between the two loci in R .

REDUCTION OF THE LINKAGE DISEQUILIBRIA

In this section, we demonstrate the rapid reduction of the linkage disequilibria d_i to $O(s)$. This result is an immediate consequence of the following useful bound. For every subset of loci $R \subseteq N$ and $t = 0, 1, 2, \dots$, the linkage disequilibria satisfy

$$|d_{i_R}^{(R)}(t)| \leq \alpha_{i_R}^{(R)} \lambda_R^t + \beta_{i_R}^{(R)} s, \quad (37)$$

where $\alpha_{i_R}^{(R)}$ and $\beta_{i_R}^{(R)}$ denote constants independent of s and t , and λ_R designates the complement of the smallest two-locus recombination frequency c_{kl} in R .

We note in passing that if there is no selection ($s = 0$), then (37) establishes global convergence of the n -locus system to linkage equilibrium at a rate no slower than λ_N^t , where

$$\lambda_N = 1 - c_{\min}. \quad (38)$$

Thus, our proof provides an alternative to the analyses of GEIRINGER (1944), BENNETT (1954), REIERSÖL (1962), and LYUBICH (1971).

We define $\alpha_i = \alpha_i^{(N)}$ and $\alpha^{(N)} = \max_i \alpha_i$, and let t_1 represent the shortest time such that $\alpha^{(N)} \lambda_N^{t_1} \leq s$. Then (37) implies

$$d_i(t) = O(s), \quad t \geq t_1 \sim (\ln s)/\ln \lambda_N, \quad (39)$$

as $s \rightarrow 0$. If $c_{\min} \ll 1$, we have the approximation $t_1 \approx -c_{\min}^{-1} \ln s$, and then t_1 may be considerably longer than the short time $-\ln s$, which will usually not exceed 5 or 10 generations. If the population is initially in linkage equilibrium, i.e., $d_i(0) = 0$ for every i , then $\alpha_{i_R}^{(R)} = 0$ for every R and i_R , so $t_1 = 0$. Therefore, t_1 should be conservative if the population is initially close to linkage equilibrium: $|d_i(0)| \ll 1$ for every i . From (24) we see that the total gene-frequency change during the time t_1 is very small, of order $st_1 \sim (s \ln s)/\ln \lambda_N$.

Applying (39) to (30) and (31), we conclude

$$\Delta p_i = O(s), \quad t \geq t_1, \quad (40)$$

$$D_i(t) = O(s), \quad t \geq t_1. \quad (41)$$

For two loci, (41) was derived in NAGYLAKI (1976); in this case, its equivalence to (39) is obvious from (36).

We shall prove (37) by induction, starting with two embedded loci and then generalizing to an arbitrary number.

Two loci: Iterating (35a) leads to

$$d_{iR}^{(R)}(t) = d_{iR}^{(R)}(0)\gamma_R^t + s \sum_{\tau=0}^{t-1} \gamma_R^\tau f_{iR}^{(R)}[\mathbf{p}(t-\tau-1), \quad t-\tau-1], \quad (42)$$

where the sum is absent for $t = 0$. Taking absolute values in (42) and substituting (35b), we get

$$|d_{iR}^{(R)}(t)| \leq |d_{iR}^{(R)}(0)| \gamma_R^t + s \rho_{iR}^{(R)} \left(\frac{1 - \gamma_R^t}{1 - \gamma_R} \right) \leq |d_{iR}^{(R)}(0)| \gamma_R^t + \frac{s \rho_{iR}^{(R)}}{1 - \gamma_R}. \quad (43)$$

For two loci, $\lambda_R = \gamma_R$, so (43) establishes (37).

Multiple loci: We show now that if (37) holds for every proper subset $R \subset N$, then it holds for N .

To estimate $g_i(p_i)$, note first from (17) that

$$|q_{iR}^{(R)}(t)| \leq 1, \quad |d_{iR}^{(R)}(t)| \leq 1. \quad (44)$$

A glance at (7) to (10) confirms the obvious, that $c_{\min} < c_{\text{tot}}$, whence (34b) and (38) yield

$$\lambda_N = \max_R \lambda_R > \gamma_N. \quad (45)$$

Equations 34d and 44 reveal that

$$|g_i(p_i)| \leq \sum_I c_I [2 |d_{iI}^{(I)}| + |d_{iI}^{(I)}|]. \quad (46)$$

Using (37) for $I \subset N$ and $J \subset N$ and then replacing λ_I and λ_J by λ_N leads to

$$|g_i(p_i)| \leq a_i \lambda_N^t + b_i s, \quad (47a)$$

where

$$a_i = \sum_I c_I (2\alpha_{iI}^{(I)} + \alpha_{iI}^{(I)}), \quad b_i = \sum_I c_I (2\beta_{iI}^{(I)} + \beta_{iI}^{(I)}). \quad (47b)$$

Substituting (34c) and (47a) into (34a), we obtain

$$|d'_i| \leq \gamma_N |d_i| + \alpha_i \lambda_N^t + B_i s, \quad (48)$$

where $B_i = \rho_i + b_i$.

Now consider $\delta_i(t)$ such that $\delta_i(0) = |d_i(0)|$ and

$$\delta'_i = \gamma_N \delta_i + \alpha_i \lambda_N^t + B_i s. \quad (49)$$

If $|d_i(t)| \leq \delta_i(t)$ for some t , then (48) and (49) inform us that $|d_i(t+1)| \leq \delta_i(t+1)$, so we conclude by

induction that

$$|d_i(t)| \leq \delta_i(t) \quad (50)$$

for $t = 0, 1, \dots$

We keep in mind (45) and apply the solution (42) to (49):

$$\begin{aligned} \delta_i(t) &= \delta_i(0)\gamma_N^t + \sum_{\tau=0}^{t-1} \gamma_N^\tau (a_i \gamma_N^{t-\tau-1} + B_i s) \\ &\leq \delta_i(0)\gamma_N^t + \frac{a_i \lambda_N^t}{\lambda_N - \gamma_N} + \frac{B_i s}{1 - \gamma_N} \\ &\leq \alpha_i \lambda_N^t + \beta_i s, \end{aligned} \quad (51a)$$

where

$$\alpha_i = \delta_i(0) + \frac{a_i}{\lambda_N - \gamma_N}, \quad \beta_i = \frac{B_i}{1 - \gamma_N}. \quad (51b)$$

Therefore, (50) and (51a) imply that (37) holds for $R = N$, which completes our inductive proof.

APPROXIMATION ON THE LINKAGE-EQUILIBRIUM SURFACE

For $t \geq t_1$, according to (39), the linkage disequilibria are $O(s)$, which suggests that the population evolves approximately as if it were in linkage equilibrium, the difference between the exact gametic frequencies and those of the much simpler system on the linkage-equilibrium surface being $O(s)$. To make this precise, recall that the exact gene frequencies $p_i^{(k)}(t)$ evolve according to the complicated law (15), which depends on the gametic frequencies p_i . The gene frequencies $\pi_i^{(k)}(t)$ on the linkage-equilibrium surface evolve according to the much simpler law obtained by imposing linkage equilibrium on (15). We choose

$$\pi_i^{(k)}(t_1) = p_i^{(k)}(t_1) \quad (52)$$

for every k and i ; we shall prove that

$$p_i(t) = \pi_i(t) + O(s), \quad t_1 \leq t \leq K/s, \quad (53a)$$

as $s \rightarrow 0$, where K designates a constant and

$$\pi_i = \prod_{k=1}^n \pi_i^{(k)} \quad (53b)$$

denotes the gametic frequencies on the linkage-equilibrium surface.

If $\pi(t)$ does not necessarily converge to some equilibrium point or if $\pi(t_1)$ is on the stable manifold of an unstable equilibrium, then small perturbations may cause large deviations in its ultimate state. In this case, the restriction $t \leq K/s$ in (53) may be necessary.

The approximation (53) justifies many analyses in which linkage equilibrium is posited. We shall invoke (53) to show that the genic covariance between an arbitrary character and fitness can be approximated on the linkage-equilibrium surface, but shall not use it otherwise.

For two loci, a different proof of (53) was presented in NAGYLAKI (1977b, pp. 171–173; 1992a, pp. 179–181).

Proof of (53): From (18a) and (39) we obtain

$$p_i(t) = q_i(t) + O(s), \quad t \geq t_1. \quad (54)$$

In view of (24), we may rewrite (15) as

$$\Delta p_{i_k}^{(k)} = sH_{i_k}^{(k)}(\mathbf{p}, t); \quad (55)$$

because of our assumptions on $W_{ij}(\mathbf{p}, t)$, the uniformly bounded function $H_{i_k}^{(k)}(\mathbf{p}, t)$ is continuously differentiable with respect to \mathbf{p} . We invoke (54) to expand (55) by Taylor's theorem:

$$\Delta p_{i_k}^{(k)} = sH_{i_k}^{(k)}(\mathbf{q}, t) + s^2 h_{i_k}^{(k)}(\mathbf{q}, t), \quad t \geq t_1, \quad (56)$$

where

$$\sum_k \sum_{i_k} |h_{i_k}^{(k)}(\mathbf{q}, t)| \leq M \quad (57)$$

for some constant M independent of \mathbf{q} , t , and s . On the linkage-equilibrium surface, (55) becomes

$$\Delta \pi_{i_k}^{(k)} = sH_{i_k}^{(k)}(\boldsymbol{\pi}, t). \quad (58)$$

We put

$$x(t) = \sum_k \sum_{i_k} |p_{i_k}^{(k)}(t) - \pi_{i_k}^{(k)}(t)|, \quad (59)$$

subtract (58) from (56), sum over k and i_k , and take absolute values to derive

$$\begin{aligned} x(t+1) &\leq \sum_k \sum_{i_k} [|p_{i_k}^{(k)}(t) - \pi_{i_k}^{(k)}(t)| \\ &\quad + s |H_{i_k}^{(k)}(\mathbf{q}, t) - H_{i_k}^{(k)}(\boldsymbol{\pi}, t)| \\ &\quad + s^2 |h_{i_k}^{(k)}(\mathbf{q}, t)|]. \end{aligned} \quad (60)$$

By Taylor's theorem, since $H_{i_k}^{(k)}(\mathbf{p}, t)$ is uniformly bounded and continuously differentiable with respect to \mathbf{p} , there exist constants $L_{i_k i_l}^{(kl)}$ independent of \mathbf{q} , $\boldsymbol{\pi}$, and t such that

$$\begin{aligned} |H_{i_k}^{(k)}(\mathbf{q}, t) - H_{i_k}^{(k)}(\boldsymbol{\pi}, t)| \\ \leq \sum_l \sum_{i_l} L_{i_k i_l}^{(kl)} |p_{i_l}^{(l)}(t) - \pi_{i_l}^{(l)}(t)|. \end{aligned} \quad (61)$$

Inserting (59), (61), and (57) into (60) leads to

$$x(t+1) \leq (1 + Ls)x(t) + Ms^2, \quad t \geq t_1, \quad (62a)$$

$$x(t_1) = 0, \quad (62b)$$

where

$$L = \max_{l, i_l} \sum_k \sum_{i_k} L_{i_k i_l}^{(kl)} \quad (63)$$

and (62b) follows from (52).

Now consider $y(t)$ such that

$$y(t+1) = (1 + Ls)y(t) + Ms^2, \quad t \geq t_1, \quad (64a)$$

$$y(t_1) = 0. \quad (64b)$$

The induction argument between (49) and (50) demonstrates that

$$x(t) \leq y(t), \quad t \geq t_1. \quad (65)$$

But (64) yields

$$\begin{aligned} y(t) &= \frac{Ms}{L} [(1 + Ls)^{t-t_1} - 1] \\ &\leq \frac{Ms}{L} (1 + Ls)^t \\ &\leq \frac{Ms}{L} e^{Lst}, \quad t \geq t_1, \end{aligned} \quad (66)$$

which establishes that

$$y(t) = O(s), \quad t_1 \leq t \leq K/s. \quad (67)$$

By (65), the same holds for $x(t)$, which proves (53).

SLOW VARIATION OF THE LINKAGE DISEQUILIBRIA

Here, we posit that the explicit time dependence (if any) of the genotypic fitnesses is $O(s^2)$:

$$W_{ij}[\mathbf{p}(t), t+1] - W_{ij}[\mathbf{p}(t), t] = O(s^2) \quad (68)$$

as $s \rightarrow 0$ for every i, j , and t . The hypothesis (68) will enable us to prove that after a time $t_2 \sim 2t_1$, the linkage disequilibria are almost constant, their rate of change being $O(s^2)$. We shall derive this conclusion from the following useful bound. For every subset of loci $R \subseteq N$, the linkage disequilibria satisfy

$$|\Delta d_{i_R}^{(R)}(t)| \leq s(\kappa_{i_R}^{(R)} \lambda_R^{-t_1} + \mu_{i_R}^{(R)} s), \quad t \geq t_1, \quad (69)$$

where $\kappa_{i_R}^{(R)}$ and $\mu_{i_R}^{(R)}$ denote constants independent of s and t .

The inequality (69) corresponds to (37). In fact, it will be clear that (69) holds with t_1 replaced by $t_R \sim (\ln s)/\ln \lambda_R$ for each R . We have used t_1 for simplicity and because our interest centers on $R = N$ and $t_N \equiv t_1$.

We define $\kappa_i = \kappa_i^{(N)}$ and $\kappa^{(N)} = \max_i \kappa_i$, and let t_2 represent the shortest time such that $\kappa^{(N)} \lambda_N^{t_2-t_1} \leq s$. Then (69) implies

$$\Delta d_i = O(s^2), \quad t \geq t_2 \sim t_1 + (\ln s)(\ln \lambda_N)^{-1} \quad (70)$$

as $s \rightarrow 0$. If the population is initially in linkage equilibrium, $t_1 = 0$; otherwise, $t_1 \sim (\ln s)/\ln \lambda_N$, so $t_2 \sim 2t_1$ as $s \rightarrow 0$. The time t_2 may be considerably longer than 10 or 20 generations if $c_{\min} \ll 1$. The total gene-frequency change during the period $t_1 < t \leq t_2$ is very small, of order $s(t_2 - t_1) \sim (s \ln s)/\ln \lambda_N$, which is approximately the same as during the initial period ($t \leq t_1$).

If the gametic frequencies converge to an equilibrium point, (24) informs us that the characteristic convergence time is $t_3 \sim 1/s \gg t_2$. Then most of the gene-frequency change occurs during the period $t_2 <$

$t \leq t_3$, which, we shall show, is precisely when we expect generic increase of the mean fitness. Of course, (39) precludes substantial change in the linkage disequilibria for $t > t_1$.

Observe that (69) and (70) differ from (37) and (39), respectively, essentially by a factor s . This is not surprising: (40) suggests that if explicit time dependence is negligible [see (68)], then for $t \geq t_1$, all functions of the gametic frequencies should change at the relative rate s , as if they were functions of st rather than of t .

The estimate (70) is the precise and general statement of quasi-linkage equilibrium and will be required for the proof of the asymptotic fundamental and secondary theorems of natural selection. KIMURA (1965) was the first to argue for two diallelic loci that the linkage disequilibria varied slowly and that this was related to the fundamental theorem of natural selection. The result (70) for \bar{W}_D and the asymptotic fundamental theorem of natural selection were proved for two multiallelic loci in NAGYLAKI (1976). Provided the position effect (if any) is $O(s^2)$ [see (135)], D_i also satisfies (70) [see below (148)].

We shall prove (69) by induction, starting with two embedded loci and then generalizing to an arbitrary number.

Two loci: Because of (39), we may set

$$d_{iR}^{(R)}(t) = s\tilde{d}_{iR}^{(R)}(t), \quad t \geq t_1, \quad (71)$$

where $\tilde{d}_{iR}^{(R)}(t)$ is uniformly bounded as $s \rightarrow 0$. Substituting (71) into (35a), we obtain

$$\tilde{d}_{iR}^{(R)'} = \gamma_R \tilde{d}_{iR}^{(R)} + f_{iR}^{(R)}(\mathbf{p}, t), \quad t \geq t_1, \quad (72)$$

whence

$$(\Delta \tilde{d}_{iR}^{(R)})' = \gamma_R \Delta \tilde{d}_{iR}^{(R)} + \Delta f_{iR}^{(R)}(\mathbf{p}, t), \quad t \geq t_1. \quad (73)$$

We decompose the change in f_i into parts due to its dependence on the gametic frequencies and on time:

$$\begin{aligned} \Delta f_{iR}^{(R)}(\mathbf{p}, t) = & \{f_{iR}^{(R)}[\mathbf{p}(t+1), t+1] \\ & - f_{iR}^{(R)}[\mathbf{p}(t), t+1]\} \\ & + \{f_{iR}^{(R)}[\mathbf{p}(t), t+1] - f_{iR}^{(R)}[\mathbf{p}(t), t]\}. \end{aligned} \quad (74)$$

By Taylor's theorem and (40), the expression in the first brace is $O(s)$. Since the selection term in (35a) is $sf_{iR}^{(R)}$, we infer from (68) that the expression in the second brace is also $O(s)$. Therefore, (73) becomes

$$(\Delta \tilde{d}_{iR}^{(R)})' = \gamma_R \Delta \tilde{d}_{iR}^{(R)} + O(s), \quad t \geq t_1, \quad (75)$$

which has the form (35a). Consequently, (43) gives

$$|\Delta \tilde{d}_{iR}^{(R)}(t)| \leq |\Delta \tilde{d}_{iR}^{(R)}(t_1)| \gamma_R^{t-t_1} + \mu_{iR}^{(R)} s, \quad t \geq t_1, \quad (76)$$

for some constants $\mu_{iR}^{(R)}$. Recalling that $\lambda_R = \gamma_R$ for two loci, we see that (71) and (76) establish (69).

Multiple loci: We demonstrate now that if (69) holds for every proper subset $R \subset N$, then it holds for N .

From (34a) we obtain

$$(\Delta d_i)' = \gamma_N \Delta d_i + s \Delta f_i(\mathbf{p}, t) + \Delta g_i(p_i). \quad (77)$$

Since the argument based on (74) applies to any number of loci, there exist constants Λ_i such that

$$|\Delta f_i(\mathbf{p}, t)| \leq \Lambda_i s. \quad (78)$$

To estimate Δg_i , first we deduce from (34d)

$$\begin{aligned} \Delta g_i(p_i) = & \sum_j c_j \{ [d_{ij}^{(j)} \Delta d_{ij}^{(j)} + d_{ij}^{(j)} \Delta d_{ij}^{(j)} + (\Delta d_{ij}^{(j)}) \Delta d_{ij}^{(j)}] \\ & + [d_{ij}^{(j)} \Delta q_{ij}^{(j)} + d_{ij}^{(j)} \Delta q_{ij}^{(j)}] + [(\Delta d_{ij}^{(j)}) \Delta q_{ij}^{(j)} \\ & + (\Delta d_{ij}^{(j)}) \Delta q_{ij}^{(j)}] + [q_{ij}^{(j)} \Delta d_{ij}^{(j)} + q_{ij}^{(j)} \Delta d_{ij}^{(j)}] \}. \end{aligned} \quad (79)$$

For $t \geq t_1$, (39), (17b), and (24) inform us that the expressions in the first three brackets are $O(s^2)$, and (44) bounds the expression in the last bracket. Therefore, we have

$$|\Delta g_i(p_i)| \leq \sum_j c_j [|\Delta d_{ij}^{(j)}| + |\Delta d_{ij}^{(j)}|] + \theta_i s^2 \quad (80)$$

for $t \geq t_1$ and some constants θ_i . Employing (69) for $I \subset N$ and $J \subset N$ and then replacing λ_i and λ_j by λ_N leads to

$$|\Delta g_i(p_i)| \leq \eta_i s \lambda_N^{t-t_1} + \nu_i s^2, \quad t \geq t_1, \quad (81a)$$

where

$$\begin{aligned} \eta_i = & \sum_j c_j (\kappa_{ij}^{(j)} + \kappa_{ij}^{(j)}), \\ \nu_i = & \theta_i + \sum_j c_j (\mu_{ij}^{(j)} + \mu_{ij}^{(j)}). \end{aligned} \quad (81b)$$

Inserting (78) and (81a) into (77), we find

$$|\Delta d_i|' \leq \gamma_N |\Delta d_i| + \eta_i s \lambda_N^{t-t_1} + \Gamma_i s^2, \quad t \geq t_1, \quad (82)$$

where $\Gamma_i = \Lambda_i + \nu_i$. By virtue of (39), we may again substitute (71) for $R = N$ to obtain

$$|\Delta \tilde{d}_i|' \leq \gamma_N |\Delta \tilde{d}_i| + \eta_i \lambda_N^{t-t_1} + \Gamma_i s, \quad t \geq t_1, \quad (83)$$

which has precisely the form (48). Therefore, (50) and (51) imply that

$$|\Delta \tilde{d}_i(t)| \leq \kappa_i \lambda_N^{t-t_1} + \mu_i s, \quad t \geq t_1, \quad (84)$$

for suitable constants κ_i and μ_i . In view of (71), this establishes (69) for $R = N$, which completes our proof.

ANALYSIS OF VARIANCE AND COVARIANCE

In this section, we demonstrate that the orthogonal decomposition of the gametic variances in an arbitrary character Z and in fitness W into their genic (or additive genetic) and epistatic components simultaneously yields the corresponding decomposition of the gametic covariance of Z and W . The right-hand sides of the asymptotic fundamental and secondary theo-

rems of natural selection are generically dominated by the genic variance in W and by the genic covariance of Z and W , respectively, and many results of our analysis will be required to prove these theorems. The analysis in this section will involve dynamics only at the end, when we shall invoke (39) to show that the average excess is close to the average effect for each allele, and then use (53) to demonstrate that the analysis of variance and covariance can be approximated on the linkage-equilibrium surface, where the allelic effects at different loci are independent.

Analysis: Let $Z_{ij}(\mathbf{p}, t)$ designate the trait value of genotype ij . We define the mean values of the gamete i , the allele $A_i^{(k)}$, and the population by

$$Z_i = \sum_j Z_{ij} p_j, \quad (85a)$$

$$p_i^{(k)} Z_i^{(k)} = \sum_i^{(k)} Z_i p_i, \quad (85b)$$

$$\bar{Z} = \sum_{i,j} Z_{ij} p_i p_j. \quad (85c)$$

For the deviations from the mean \bar{Z} , we have

$$z_{ij} = Z_{ij} - \bar{Z}, \quad (86a)$$

$$z_i = Z_i - \bar{Z} = \sum_j z_{ij} p_j, \quad (86b)$$

$$p_i^{(k)} z_i^{(k)} = p_i^{(k)} (Z_i^{(k)} - \bar{Z}) = \sum_i^{(k)} z_i p_i. \quad (86c)$$

Thus, z_i and $z_i^{(k)}$ signify the average excesses for the character of the gamete i and the allele $A_i^{(k)}$, respectively.

We decompose the gametic excess z_i into an additive effect ζ_i and an epistatic deviation e_i ,

$$z_i = \zeta_i + e_i, \quad (87a)$$

and then set

$$\zeta_i = \sum_{k=1}^n \zeta_i^{(k)}, \quad (87b)$$

where $\zeta_i^{(k)}$ denotes the average effect of the allele $A_i^{(k)}$ on the character Z , determined below. Then the gametic, genic, and epistatic components of the variance in Z are

$$V_{\text{gam}} = 2 \sum_i p_i z_i^2, \quad (88a)$$

$$V_g = 2 \sum_i p_i \zeta_i^2, \quad (88b)$$

$$V_e = 2 \sum_i p_i e_i^2. \quad (88c)$$

We eliminate e_i from (88c) with the aid of (87) and minimize V_e with respect to $\zeta_i^{(k)}$. This leads easily to [cf. EWENS (1979), pp. 217–218]

$$\sum_i^{(k)} p_i e_i = 0 \quad (89)$$

for every k and i_k . Substituting (87a) into (86c) and

using (89) yields

$$p_{i_k}^{(k)} z_{i_k}^{(k)} = \sum_i^{(k)} p_i \zeta_i. \quad (90)$$

In (90), there is one equation for each k and i_k . The unknown allelic effects $\zeta_{i_k}^{(k)}$ appear only in the sums ζ_i ; summing (89) over i_k and using (86b) and (86a), we infer

$$\sum_i p_i \zeta_i = 0. \quad (91)$$

Let us momentarily translate the average effects according to

$$\zeta_{i_k}^{(k)} = \tilde{\zeta}_{i_k}^{(k)} + \sigma_k$$

and observe that the choice

$$\sigma_k = \sum_{i_k} p_{i_k}^{(k)} \zeta_{i_k}^{(k)}$$

gives

$$\sum_{i_k} p_{i_k}^{(k)} \tilde{\zeta}_{i_k}^{(k)} = 0.$$

Furthermore, (91) and (87b) reveal that

$$\sum_k \sigma_k = 0,$$

whence $\zeta_i = \tilde{\zeta}_i$. Therefore, we may simply assume

$$\sum_{i_k} p_{i_k}^{(k)} \zeta_{i_k}^{(k)} = 0 \quad (92)$$

for each k .

Appealing successively to (88b), (87b), and (90), we deduce

$$\begin{aligned} V_g &= 2 \sum_i p_i \zeta_i \sum_k \zeta_{i_k}^{(k)} \\ &= 2 \sum_k \sum_{i_k} \zeta_{i_k}^{(k)} \sum_i^{(k)} p_i \zeta_i \\ &= 2 \sum_k \sum_{i_k} p_{i_k}^{(k)} \zeta_{i_k}^{(k)} z_{i_k}^{(k)}. \end{aligned} \quad (93)$$

From (88a), (87a), (88b), (88c), (87b), and (89), we derive

$$\begin{aligned} V_{\text{gam}} &= 2 \sum_i p_i (\zeta_i^2 + e_i^2 + 2\zeta_i e_i) \\ &= V_g + V_e + 4 \sum_k \sum_{i_k} \zeta_{i_k}^{(k)} \sum_i^{(k)} p_i e_i \\ &= V_g + V_e. \end{aligned} \quad (94)$$

Since the gametes are combined in generalized Hardy-Weinberg proportions, therefore V_{gam} is the additive component of the total genetic variance V in the least-squares decomposition

$$V = V_{\text{gam}} + V_{\text{gam}, \text{dom}}, \quad (95)$$

where $V_{\text{gam}, \text{dom}}$ represents the variance that arises from nonadditivity of the gametic effects. Consequently, V_g is indeed the additive component of V (KIMURA 1965; EWENS 1979, pp. 217–218; NAGYLAKI 1992a, p. 182).

Since our analysis is entirely within gametes, dominance deviations cannot enter. Therefore, the epistatic deviation e_i and variance V_e involve only the interaction of additive effects, into which we proceed to decompose them. The reader may find some aspects of BULMER's (1980, pp. 46–51) panmictic analysis of the diploid variance helpful.

Suppose here that $R \subseteq N$ contains at least two loci and put

$$e_i = \sum_{R \subseteq N} x_{iR}^{(R)}, \quad (96)$$

where $x_{iR}^{(R)}$ designates the effect of the interaction of all the loci in R . The corresponding components of the epistatic variance V_e are

$$V_e^{(R)} = 2 \sum_{iR} p_i [x_{iR}^{(R)}]^2. \quad (97)$$

Successive minimization of these variance components for sets R of $n, n-1, \dots, 3$ loci leads to

$$\sum_i p_i x_{iR}^{(R)} = 0 \quad \text{for every } k \in R. \quad (98)$$

Substituting (96) into (88c) yields

$$V_e = 2 \sum_{R, S \subseteq N} \sum_i p_i x_{iR}^{(R)} x_{iS}^{(S)}. \quad (99)$$

If $R \neq S$, then there exists $k \in R$ such that $k \notin S$, or $k \in S$ such that $k \notin R$. Together with (98), (16), and (97), this fact reduces (99) to orthogonal form:

$$V_e = 2 \sum_{R \subseteq N} \sum_i p_i [x_{iR}^{(R)}]^2 = \sum_{R \subseteq N} V_e^{(R)}. \quad (100)$$

We remark that (96) and (98) enable us to confirm (89):

$$\sum_i^{(k)} p_i e_i = \sum_{R \subseteq N} \sum_i^{(k)} p_i x_{iR}^{(R)} = 0,$$

in which the inner sum vanishes because R contains at least two loci.

We define the mean deviation $e_{iR}^{(R)}$ by

$$p_{iR}^{(R)} e_{iR}^{(R)} = \sum_{iQ} p_i e_i, \quad (101)$$

in which, as always, $Q = N - R$. Inserting (96) into (101) and using (98) and (16), we find

$$\begin{aligned} p_{iR}^{(R)} e_{iR}^{(R)} &= \sum_{S \subseteq N} \sum_{iQ} p_i x_{iS}^{(S)} \\ &= \sum_{S \subseteq R} \sum_{iQ} p_i x_{iS}^{(S)} \\ &= p_{iR}^{(R)} \sum_{S \subseteq R} x_{iS}^{(S)}, \end{aligned}$$

whence

$$e_{iR}^{(R)} = \sum_{S \subseteq R} x_{iS}^{(S)}. \quad (102)$$

Thus, $e_{iR}^{(R)}$ is the sum of the effects of the interactions

in all the subsets of R . We can rewrite (102) as

$$x_{iR}^{(R)} = e_{iR}^{(R)} - \sum_{S \subset R} x_{iS}^{(S)}, \quad (103)$$

from which, starting with $x_{iR}^{(R)} = e_{iR}^{(R)}$ for subsets of two loci, $x_{iR}^{(R)}$ can be determined recursively as a linear function of the known deviations $e_{iS}^{(S)}$ with $S \subset R$.

We turn now to the analysis of covariance. Corresponding to (87), we decompose the gametic fitness excess w_i into an additive effect ξ_i and an epistatic deviation u_i ,

$$w_i = \xi_i + u_i, \quad (104a)$$

and then set

$$\xi_i = \sum_{k=1}^n \xi_i^{(k)}, \quad (104b)$$

where $\xi_i^{(k)}$ denotes the average effect of $A_i^{(k)}$ on fitness. We define the gametic, genic, and epistatic covariances of Z and W as

$$C_{\text{gam}} = 2 \sum_i p_i z_i w_i, \quad (105a)$$

$$C_g = 2 \sum_i p_i \xi_i \xi_i, \quad (105b)$$

$$C_e = 2 \sum_i p_i e_i u_i. \quad (105c)$$

Appealing successively to (105a), (87a), (104a), (87b), (104b), (89), and the analog of (89) for u_i , we find

$$\begin{aligned} C_{\text{gam}} &= 2 \sum_i p_i (\xi_i \xi_i + e_i u_i + \xi_i u_i + e_i \xi_i) \\ &= C_g + C_e + 2 \sum_k \sum_{i_k} \left(\xi_{i_k}^{(k)} \sum_i p_i u_i + \xi_{i_k}^{(k)} \sum_i p_i e_i \right) \\ &= C_g + C_e. \end{aligned} \quad (106)$$

To justify the term genic covariance, we express C_g purely in terms of allelic variables. Consecutive use of (105b), (87b), and the analog of (90) for fitness produces

$$\begin{aligned} C_g &= 2 \sum_k \sum_{i_k} \xi_{i_k}^{(k)} \sum_i p_i \xi_i \\ &= 2 \sum_k \sum_{i_k} p_{i_k}^{(k)} \xi_{i_k}^{(k)} w_{i_k}^{(k)}, \end{aligned} \quad (107a)$$

which is the covariance of the average effect on the character ($\xi_{i_k}^{(k)}$) and the average excess for fitness ($w_{i_k}^{(k)}$) of every allele that affects the character. Observe the loss of symmetry between Z and W due to the decomposition of (105b) into allelic effects. However, applying (104b) and (90) to (105b) yields

$$C_g = 2 \sum_k \sum_{i_k} p_{i_k}^{(k)} z_{i_k}^{(k)} \xi_{i_k}^{(k)} \quad (107b)$$

as an alternative to (107a), which shows that the loss of symmetry is only apparent. Since, according to (15),

it is $w_{i_k}^{(k)}$ rather than $\xi_{i_k}^{(k)}$ that controls gene-frequency change, (107a) is more natural and useful than (107b).

To decompose the epistatic covariance C_e , we set

$$u_i = \sum_{R \subseteq N} v_{i_R}^{(R)}, \quad (108)$$

where $v_{i_R}^{(R)}$ designates the effect on fitness of the interaction of all the loci in R . We define the corresponding components of C_e as

$$C_e^{(R)} = 2 \sum_{i_R} p_{i_R}^{(R)} x_{i_R}^{(R)} v_{i_R}^{(R)}. \quad (109)$$

Substituting (96) and (108) into (105c), utilizing (98) and its analog for $v_{i_R}^{(R)}$ as in the proof of (100), and employing (16) and (109), we deduce

$$C_e = \sum_{R \subseteq N} C_e^{(R)}. \quad (110)$$

Next, we prove that the average excess $z_{i_k}^{(k)}$ and the average effect $\xi_{i_k}^{(k)}$ differ only because of linkage disequilibrium. To see this, in successive lines we invoke (90) and (87b); (4); (16) and (17); (92); and (39):

$$\begin{aligned} p_{i_k}^{(k)} z_{i_k}^{(k)} &= \sum_i^{(k)} p_i \left(\xi_{i_k}^{(k)} + \sum_{l: l \neq k} \zeta_{i_l}^{(l)} \right) \\ &= p_{i_k}^{(k)} \xi_{i_k}^{(k)} + \sum_{l: l \neq k} \sum_{i_l} \zeta_{i_l}^{(l)} \sum_i^{(k,l)} p_i \\ &= p_{i_k}^{(k)} \xi_{i_k}^{(k)} + \sum_{l: l \neq k} \sum_{i_l} \zeta_{i_l}^{(l)} (p_{i_k}^{(k)} p_{i_l}^{(l)} + d_{i_k i_l}^{(k,l)}) \\ &= p_{i_k}^{(k)} \xi_{i_k}^{(k)} + \sum_{l: l \neq k} \sum_{i_l} d_{i_k i_l}^{(k,l)} \zeta_{i_l}^{(l)} \end{aligned} \quad (111)$$

$$= p_{i_k}^{(k)} \xi_{i_k}^{(k)} + O(s), \quad t \geq t_1; \quad (112)$$

the sum in the second line above is over all components of the vector i except i_k and i_l . For fitness, (111) holds with $z_{i_k}^{(k)}$ and $\xi_{i_k}^{(k)}$ replaced by $w_{i_k}^{(k)}$ and $\xi_{i_k}^{(k)}$, respectively, but, on account of (22), the approximation (112) becomes

$$p_{i_k}^{(k)} w_{i_k}^{(k)} = p_{i_k}^{(k)} \xi_{i_k}^{(k)} + O(s^2), \quad t \geq t_1. \quad (113)$$

Finally, we observe that (53) and Taylor's theorem enable us to approximate the variances and covariances on the linkage-equilibrium surface. For the variance components of Z , (88) gives

$$V_{\text{gam}}(\mathbf{p}, t) = V_{\text{gam}}(\boldsymbol{\pi}, t) + O(s), \quad t_1 \leq t \leq K/s, \quad (114)$$

and analogous approximations for V_g and V_e . From (105) we get

$$C_{\text{gam}}(\mathbf{p}, t) = C_{\text{gam}}(\boldsymbol{\pi}, t) + O(s^2), \quad t_1 \leq t \leq K/s, \quad (115)$$

and analogous formulas for C_g and C_e . For the variance components of W , (88) yields

$$V_{\text{gam}}(\mathbf{p}, t) = V_{\text{gam}}(\boldsymbol{\pi}, t) + O(s^3), \quad t_1 \leq t \leq K/s, \quad (116)$$

and similar expressions for V_g and V_e . Of course, the leading terms in (114), (115), and (116) are $O(1)$, $O(s)$, and $O(s^2)$, respectively.

On the linkage-equilibrium surface, we have a simple expansion of the epistatic variance V_e in terms of the components V_{A^r} , of the total genetic variance V associated with all sets of r additive effects on Z (NAGYLAKI 1992b):

$$V_e(\boldsymbol{\pi}, t) = \sum_{r=2}^n 2^{1-r} V_{A^r}(\boldsymbol{\pi}, t). \quad (117)$$

This formula generally does not hold if there is linkage disequilibrium.

EPISTASIS AND LINKAGE DISEQUILIBRIUM

As stated in the introduction, if fitnesses are additive between loci, the gametic frequencies always converge to a stationary point in linkage equilibrium (KARLIN and FELDMAN 1970; KARLIN 1978; KUN and LYUBICH 1979). Therefore, it ought to be possible to relate the linkage disequilibria to the epistatic deviations in fitness that maintain them (*cf.* FELSENSTEIN 1965; LANGLEY and CROW 1974; HASTINGS 1985, 1986; BARTON 1986). This is accomplished here.

The desired relations follow from the equation

$$\Delta d_i = \bar{W}^{-1} p_i u_i - D_i + O(s^2), \quad t \geq t_1, \quad (118)$$

as $s \rightarrow 0$, where the linkage disequilibria d_i and D_i are defined by (18) and (11b), respectively, and u_i represents the epistatic deviation in fitness, defined by (104). Recalling (105c), from (118) we obtain immediately

$$2 \sum_i e_i \Delta d_i = \bar{W}^{-1} C_e - 2 \sum_i e_i D_i + O(s^2), \quad t \geq t_1, \quad (119)$$

which will provide a simple proof of the asymptotic fundamental and secondary theorems of natural selection.

For two multiallelic loci, (118) and (119) hold also with D_i instead of d_i on the left-hand side (NAGYLAKI 1976).

Suppose the explicit time dependence (if any) of the genotypic fitnesses is sufficiently weak to satisfy (68). Then (118) and (70) give

$$D_i(t) = \bar{W}^{-1} p_i u_i + O(s^2), \quad t \geq t_2. \quad (120)$$

Several interesting conclusions follow from this remarkably simple relation. If epistasis is weak in the sense that $u_i = O(s^2)$, as is the case for multiplicative fitnesses, then $D_i(t) = O(s^2)$ for $t \geq t_2$, which is much stronger than (41). This result was proved for two multiallelic loci in NAGYLAKI (1977a). If, on the con-

trary, u_i is comparable to s in magnitude, then $D_i(t) \approx \bar{W}^{-1} p_i u_i$ for $t \geq t_2$. Thus, in this case, D_i has the same sign and order of magnitude as the epistatic deviation in fitness.

From (119) and (70) we obtain

$$C_e = 2\bar{W} \sum_i e_i D_i + O(s^2), \quad t \geq t_2; \quad (121)$$

for the epistatic variance in fitness, (121) becomes

$$V_e = 2\bar{W} \sum_i u_i D_i + O(s^3), \quad t \geq t_2. \quad (122)$$

According to (108), u_i is the sum of the effects of the fitness interactions in all subsets of loci. To illustrate this point, we define for the subset R of at least two loci the linkage disequilibria $D_{i_R}^{(R)}$ and mean epistatic deviations $u_{i_R}^{(R)}$ by

$$D_{i_R}^{(R)} = \sum_{i_Q} D_{i_Q}, \quad (123)$$

$$p_{i_R}^{(R)} u_{i_R}^{(R)} = \sum_{i_Q} p_{i_Q} u_{i_Q}, \quad (124)$$

where Q denotes the complement of R . Then (120) yields

$$D_{i_R}^{(R)}(t) = \bar{W}^{-1} p_{i_R}^{(R)} u_{i_R}^{(R)} + O(s^2), \quad t \geq t_2. \quad (125)$$

Recalling (96), (102), and (108), we get

$$u_{i_R}^{(R)} = \sum_{S \subseteq R} v_{i_S}^{(S)}. \quad (126)$$

Thus, for each subset R , the linkage disequilibrium $D_{i_R}^{(R)}$ is proportional to the sum of the effects of the fitness interactions in all subsets of R . The interactions in all subsets $S \subseteq R$ contribute to $D_{i_R}^{(R)}$ because, according to (123), all the linkage disequilibria $D_{i_S}^{(S)}$ are "contained" in $D_{i_R}^{(R)}$.

To decompose (125) so that only $v_{i_R}^{(R)}$ appears on the right-hand side, we generalize SLATKIN's (1972) linkage disequilibria to multiple alleles. We define for gametes the allelic indicator random variables $\chi_{i_k}^{(k)}$ by

$$\chi_{i_k}^{(k)} = \begin{cases} 1 & \text{if } A_{i_k}^{(k)} \text{ is the allele at locus } k, \\ 0 & \text{otherwise,} \end{cases} \quad (127)$$

and write \mathcal{E} for the expectation over the gametic frequencies p_i . Our new linkage disequilibria are the moments

$$\Delta_{i_R}^{(R)} = \mathcal{E} \left[\prod_{k \in R} (\chi_{i_k}^{(k)} - p_{i_k}^{(k)}) \right]. \quad (128)$$

Observe that

$$\sum_{i_k} \chi_{i_k}^{(k)} = 1 \quad (129)$$

for every k , which demonstrates immediately that

$$\sum_{i_k} \Delta_{i_R}^{(R)} = 0 \quad (130)$$

for every $k \in R$, in contrast to (123). The work of HASTINGS (1986) and BARTON (1986) suggests that $\Delta_{i_R}^{(R)}$ might depend only on $v_{i_R}^{(R)}$ for weak selection.

Before presenting our result, we must describe recombination in the subset R . We define the set of sets

$$N_K^{(R)} = \{M: K \subseteq M \subseteq N \text{ and } R - K \subseteq N - M\} \quad (131)$$

for $K \subseteq R$. We extend our definition of the recombination frequencies c_I in the biologically natural manner ($1 \in I \subseteq N$):

$$c_{N-I} = c_I. \quad (132)$$

Then the probability of reassociation of the genes at the loci in K , inherited from one parent, with the genes at the loci in $R - K$, inherited from the other, is

$$c_K^{(R)} = \sum_{M \subseteq N_K^{(R)}} c_M. \quad (133)$$

These are, of course, marginal probabilities. If r_1 is the smallest element of R , we can write the total probability of recombination among the loci in R as

$$c_{\text{tot}}^{(R)} = \sum_{K: r_1 \in K \subseteq R} c_K^{(R)} = \frac{1}{2} \sum_{K \subseteq R} c_K^{(R)}. \quad (134)$$

Observe that (131), (133), and (134) generalize (8), (9), and (7), respectively; the notation here takes advantage of (132).

We assume that the position effect (if any) is $O(s^2)$:

$$W_{ij} - W_{i|Jj, j|I} = O(s^2) \quad (135)$$

for every i, j , and I , where $J = N - I$. We shall establish that

$$\Delta_{i_R}^{(R)}(t) = (c_{\text{tot}}^{(R)} \bar{W})^{-1} p_{i_R}^{(R)} v_{i_R}^{(R)} + O(s^2), \quad t \geq t_2, \quad (136)$$

for every subset $R \subseteq N$. This surprisingly simple result seems to be quite difficult to prove, but our proof is highly informative: we shall express $\Delta_{i_R}^{(R)}$ and $D_{i_R}^{(R)}$ in terms of $d_{i_S}^{(S)}$ for $S \subseteq R$ [see (142) and (148)] and then $\Delta_{i_R}^{(R)}$ in terms of $D_{i_S}^{(S)}$ [see (149)]. Note that the form of (136) differs from that of (125) only by the factor $c_{\text{tot}}^{(R)}$. Such a difference was to be expected because (11b) shows that the explicit dependence of $D_{i_R}^{(R)}$ on the recombination frequencies is linear, whereas the definition (128) of $\Delta_{i_R}^{(R)}$ is independent of them.

From (134) we infer that if $v_{i_R}^{(R)} = O(s^2)$ for some R , then $\Delta_{i_R}^{(R)}(t) = O(s^2)$ for $t \geq t_2$. For multiplicative fitnesses, both the assumption and conclusion hold for every R . If, on the contrary, $v_{i_R}^{(R)}$ is comparable to s in magnitude, then the error term in (136) may be neglected, and therefore $\Delta_{i_R}^{(R)}$ has the same sign and order of magnitude as $v_{i_R}^{(R)}$.

Proof of (118): In successive lines, we appeal to (18a); (11a), (18b), and (24); (104), (18a), and (15);

(22) and (39); and (113):

$$\begin{aligned}
 \Delta d_i &= \Delta p_i - \Delta q_i \\
 &= \bar{W}^{-1} p_i w_i - D_i - \sum_k (q_i / p_{i_k}^{(k)}) \Delta p_{i_k}^{(k)} + O(s^2) \\
 &= \bar{W}^{-1} p_i u_i - D_i + \bar{W}^{-1} p_i \sum_k \xi_{i_k}^{(k)} \\
 &\quad - \bar{W}^{-1} (p_i - d_i) \sum_k w_{i_k}^{(k)} + O(s^2) \\
 &= \bar{W}^{-1} p_i u_i - D_i + \bar{W}^{-1} p_i \sum_k (\xi_{i_k}^{(k)} - w_{i_k}^{(k)}) + O(s^2) \\
 &= \bar{W}^{-1} p_i u_i - D_i + O(s^2), \tag{137}
 \end{aligned}$$

where the last two lines hold for $t \geq t_1$.

Proof of (136): The proof falls naturally into several sections.

First, we derive from (128) an explicit formula for $\Delta_{i_R}^{(R)}$. We define the gametic indicator random variables

$$\chi_{i_K}^{(K)} = \prod_{k \in K} \chi_{i_k}^{(k)} \tag{138}$$

$$= \begin{cases} 1 & \text{if } i_K \text{ is the gamete in } K, \\ 0 & \text{otherwise.} \end{cases} \tag{139}$$

Setting $L = R - K$ and denoting the number of elements of L by $l = |L|$, from (128), (139), and (17) we find

$$\begin{aligned}
 \Delta_{i_R}^{(R)} &= \sum_{K \subseteq R} (-1)^l p_{i_K}^{(K)} q_{i_L}^{(L)} \\
 &= a_R q_{i_R}^{(R)} + \sum_{K \subset R} (-1)^l d_{i_K}^{(K)} q_{i_L}^{(L)}, \tag{140a}
 \end{aligned}$$

where

$$a_R = \sum_{K \subseteq R} (-1)^l = \sum_{L \subseteq R} (-1)^l. \tag{140b}$$

For fixed l and $|R| = r$, the set L can be chosen in $\binom{r}{l}$ ways, so

$$a_R = \sum_{l=0}^r \binom{r}{l} (-1)^l = (1 - 1)^r = 0, \tag{141}$$

and therefore

$$\Delta_{i_R}^{(R)} = \sum_{K \subseteq R} (-1)^l q_{i_L}^{(L)} d_{i_K}^{(K)}. \tag{142}$$

As a check, one can show that (142) satisfies (130). We note parenthetically that (142), (17b), (24), (39), and (70) imply that $\Delta_{i_R}^{(R)}$ satisfies (70) for every $R \subseteq N$.

Second, we approximate $D_{i_R}^{(R)}$ for weak selection. We rewrite (11b) as

$$D_i = \sum_j \sum_t c_t Y_{ij}^{(t)}, \tag{143a}$$

where $1 \in I \subset N$ and

$$Y_{ij}^{(t)} = \bar{W}^{-1} (W_{ij} p_i p_j - W_{i_{ij} j_{ij}} p_{i_{ij}} p_{j_{ij}}). \tag{143b}$$

Inserting (18a), (39), and (17b) into (143b) leads to

$$\begin{aligned}
 Y_{ij}^{(t)} &= \bar{W}^{-1} (W_{ij} - W_{i_{ij} j_{ij}}) q_i q_j \\
 &\quad + \bar{W}^{-1} [W_{ij} (q_i d_j + q_j d_i) \\
 &\quad - W_{i_{ij} j_{ij}} (q_{i_{ij}} d_{j_{ij}} + q_{j_{ij}} d_{i_{ij}})] \\
 &\quad + O(s^2), \quad t \geq t_1. \tag{144}
 \end{aligned}$$

We impose (135) and use (21b) to simplify (44) to

$$Y_{ij}^{(t)} = q_i d_j + q_j d_i - q_{i_{ij}} d_{j_{ij}} - q_{j_{ij}} d_{i_{ij}} + O(s^2), \quad t \geq t_1, \tag{145}$$

and substitute this into (143a) and invoke (16) and (17) to deduce

$$D_i(t) = \sum_t c_t (d_i - q_{i_t}^{(t)} d_{i_t}^{(t)} - q_{i_t}^{(t)} d_{i_t}^{(t)}) + O(s^2), \quad t \geq t_1. \tag{146}$$

With the aid of (7) and (132), we immediately reduce (146) to

$$D_i(t) = c_{\text{tot}} d_i - \sum_{S \subseteq N} c_S q_{i_T}^{(T)} d_{i_S}^{(S)} + O(s^2), \quad t \geq t_1, \tag{147}$$

where $T = N - S$.

Now it is obvious that

$$D_{i_R}^{(R)}(t) = c_{\text{tot}}^{(R)} d_{i_R}^{(R)} - \sum_{K \subseteq R} c_K^{(R)} q_{i_L}^{(L)} d_{i_K}^{(K)} + O(s^2), \quad t \geq t_1, \tag{148}$$

in which $L = R - K$. Indeed, one can verify (148) by utilizing (123), (7), (133), and (134) in (147).

We note in passing that (148), (17b), (24), (39), and (70) imply that $D_{i_R}^{(R)}$ satisfies (70) for every $R \subseteq N$.

Since we know $D_{i_R}^{(R)}$ from (125), our third task is to express $\Delta_{i_R}^{(R)}$ in terms of $D_{i_S}^{(S)}$ for $S \subseteq R$. For 2, 3, and 4 embedded loci, (148) was solved for $d_{i_R}^{(R)}$ in terms of $D_{i_S}^{(S)}$, and the result was substituted into (142). These increasingly tedious calculations suggest the surprisingly simple formula

$$c_{\text{tot}}^{(R)} \Delta_{i_R}^{(R)}(t) = \sum_{K \subseteq R} (-1)^l q_{i_L}^{(L)} D_{i_K}^{(K)} + O(s^2), \quad t \geq t_1. \tag{149}$$

To simplify the notation, we shall prove (149) for $R = N$; on account of the invariant form of our equations, this is clearly sufficient. Thus, we must demonstrate that

$$c_{\text{tot}} \Delta_i(t) = \sum_{S \subseteq N} (-1)^r q_{i_T}^{(T)} D_{i_S}^{(S)} + O(s^2), \quad t \geq t_1, \tag{150}$$

in which $T = N - S$ and $\tau = |T|$. Inserting (148) into the right-hand side of (150) suggests the definition

$$F_i = \sum_{S \subseteq N} (-1)^r q_{i_T}^{(T)} \left[c_{\text{tot}}^{(S)} d_{i_S}^{(S)} - \sum_{K \subseteq S} c_K^{(S)} q_{i_L}^{(L)} d_{i_K}^{(K)} \right], \tag{151}$$

where now $L = S - K$. We shall establish (150) by

proving the identity

$$F_i = c_{\text{tot}} \Delta_i, \quad (152)$$

where Δ_i is given by (142).

We segregate the term $S = N$ in the single sum in (151), set $M = N - K$, observe that $q_{i_T}^{(T)} q_{i_L}^{(L)} = q_{i_M}^{(M)}$, and interchange the summations in the double sum:

$$F_i = c_{\text{tot}} d_i + \sum_{S \subseteq N} (-1)^\tau c_{\text{tot}}^{(S)} q_{i_T}^{(T)} d_{i_S}^{(S)} - \sum_{K \subseteq N} \sum_{S: K \subseteq S \subseteq N} (-1)^\tau c_K^{(S)} q_{i_M}^{(M)} d_{i_K}^{(K)}. \quad (153)$$

We put $\kappa = |K|$ and $\sigma = |S|$ and interchange the names of the sets K and S in the double sum. Then $(-1)^\tau$ becomes

$$(-1)^{n-\kappa} = (-1)^{n-\sigma} (-1)^{\sigma-\kappa} = (-1)^\tau (-1)^{\kappa-\sigma},$$

where we have retained the definition $\tau = n - \sigma$. Thus, we transform (153) to

$$F_i = c_{\text{tot}} d_i + \sum_{S \subseteq N} (-1)^\tau \varphi_S q_{i_T}^{(T)} d_{i_S}^{(S)}, \quad (154a)$$

in which

$$\varphi_S = c_{\text{tot}}^{(S)} + \psi_S, \quad (154b)$$

$$\psi_S = - \sum_{K: S \subseteq K \subseteq N} (-1)^{\kappa-\sigma} c_S^{(K)}, \quad (154c)$$

and $T = N - S$, as above.

Substituting (133) into (154c), appealing to (131), and interchanging summations yields

$$\psi_S = - \sum_{M: S \subseteq M \subseteq N} c_M b_{S,M}, \quad (155a)$$

where

$$b_{S,M} = \sum_{\substack{K: S \subseteq K \subseteq N \\ K-S \subseteq N-M}} (-1)^{\kappa-\sigma}. \quad (155b)$$

Given κ , we can choose $L = K - S \subseteq N - M$ in $\binom{n-m}{l}$ ways, where $m = |M|$. Furthermore, the restrictions $S \subseteq K$ and $M \subseteq N$ imply respectively that $l > 0$ and $m < n$. Consequently, we have

$$b_{S,M} = \sum_{l=1}^{n-m} \binom{n-m}{l} (-1)^l = (1-1)^{n-m} - 1 = -1, \quad (156)$$

so (155a) reduces to

$$\psi_S = \sum_{M: S \subseteq M \subseteq N} c_M, \quad (157)$$

which is the total probability of recombination events in N such that the loci in S do not recombine. Therefore, it is evident from (154b) that $\varphi_S = c_{\text{tot}}$, and this can be proved by manipulation of (133) and (134). Returning to (154a), we see that

$$F_i = c_{\text{tot}} \sum_{S \subseteq N} (-1)^\tau q_{i_T}^{(T)} d_{i_S}^{(S)}, \quad (158)$$

a glance at (142) informs us that (158) is identical to (152).

We are finally ready to prove (136). Note first that (17a) and (39) enable us to rewrite (125) as

$$D_{i_R}^{(R)}(t) = \bar{W}^{-1} q_{i_R}^{(R)} u_{i_R}^{(R)} + O(s^2), \quad t \geq t_2. \quad (159)$$

Substituting (159) into (149) leads to

$$c_{\text{tot}}^{(R)} \bar{W} \Delta_{i_R}^{(R)}(t) = q_{i_R}^{(R)} G_{i_R}^{(R)} + O(s^2), \quad t \geq t_2, \quad (160a)$$

where

$$G_{i_R}^{(R)} = \sum_{K \subseteq R} (-1)^l u_{i_K}^{(K)}, \quad (160b)$$

in which $L = R - K$ and $l = |L|$.

We substitute (126) into (160b) and interchange summations:

$$G_{i_R}^{(R)} = \sum_{S \subseteq R} v_{i_S}^{(S)} \sum_{K: S \subseteq K \subseteq R} (-1)^{r-\kappa} = v_{i_R}^{(R)} + \sum_{S \subseteq R} \omega_{R,S} v_{i_S}^{(S)}, \quad (161a)$$

where

$$\omega_{R,S} = \sum_{K: S \subseteq K \subseteq R} (-1)^{r-\kappa}. \quad (161b)$$

Setting $\sigma = |S| < r$ and noting that, for fixed κ , we can choose K in $\binom{r-\sigma}{\kappa} = \binom{r-\sigma}{r-\kappa}$ ways, we find

$$\omega_{R,S} = \sum_{\kappa=\sigma}^r \binom{r-\sigma}{r-\kappa} (-1)^{r-\kappa} = (1-1)^{r-\sigma} = 0. \quad (162)$$

We conclude from (161a) that $G_{i_R}^{(R)} = v_{i_R}^{(R)}$, whence (160a) yields

$$c_{\text{tot}}^{(R)} \bar{W} \Delta_{i_R}^{(R)}(t) = q_{i_R}^{(R)} v_{i_R}^{(R)} + O(s^2), \quad t \geq t_2, \quad (163)$$

and (17a) and (39) demonstrate that (163) is equivalent to (136).

THE ASYMPTOTIC FUNDAMENTAL AND SECONDARY THEOREMS OF NATURAL SELECTION

The results of the previous sections enable us to prove the asymptotic fundamental and secondary theorems of natural selection. We shall derive the former as a special case of the latter. Although (176) below, together with (11), gives the rate of change $\Delta \bar{Z}$ exactly, this formula is unilluminating because of its complexity. The secondary theorem follows from the instructive approximation

$$\Delta \bar{Z} = \bar{W}^{-1} C_g + \bar{\Delta Z} + 2 \sum_i e_i \Delta d_i + O(s^2), \quad t \geq t_1, \quad (164a)$$

as $s \rightarrow 0$, where the genic covariance C_g is given by (107),

$$\bar{\Delta Z} = \sum_{i,j} (\Delta Z_{ij}) p_i' p_j' \quad (164b)$$

signifies the mean of the changes in trait genotypic values over the next generation, and e_i denotes the epistatic deviation in (87).

The result (164) depends only on the weakness of selection; there are no restrictions on frequency or time dependence.

Note that the sum in (164a) is absent if either the character is determined without epistasis ($e_i = 0$ for every i) or the linkage disequilibria are constant ($\Delta d_i = 0$ for every i). In the first case, (164) agrees with the exact formula in NAGYLAKI (1989). The second case occurs if fitnesses are multiplicative and the population is initially in linkage equilibrium, *i.e.*, if $d_i(0) = 0$ for every i , then $d_i(t) = 0$ for every i and t (MORAN 1967; KARLIN 1975; NAGYLAKI 1992a, Problems 8.2 and 8.10).

According to (115), we can approximate the genic covariance on the linkage-equilibrium surface:

$$C_g(\mathbf{p}, t) = C_g(\boldsymbol{\pi}, t) + O(s^2), \quad t_1 \leq t \leq K/s. \quad (165)$$

Thus, we could replace $C_g(\mathbf{p}, t)$ in (164a) by its linkage-equilibrium approximation $C_g(\boldsymbol{\pi}, t)$ without changing the order of the error.

If the explicit time dependence of the fitnesses (if any) is sufficiently weak to satisfy (68), then (70) holds, and hence (164a) reduces to

$$\Delta \bar{Z} = \bar{W}^{-1} C_g + \bar{\Delta Z} + O(s^2), \quad t \geq t_2. \quad (166)$$

If, in addition, the genotypic values Z_{ij} vary sufficiently slowly to satisfy

$$\Delta Z_{ij} = O(s^2) \quad (167)$$

for every i and j , we obtain the asymptotic secondary theorem of natural selection:

$$\Delta \bar{Z} = \bar{W}^{-1} C_g + O(s^2), \quad t \geq t_2, \quad (168)$$

as $s \rightarrow 0$. The weaker assumption $\Delta Z_{ij} = o(s)$ (*i.e.*, of an order smaller than s) leads to (168) with the error $o(s)$ instead of $O(s^2)$. The *absolute* error in (168) is small, though the *relative* error may be large.

Whereas for quantitative genetics our interest centers on the evolution of \bar{Z} , it is the evolution of the mean fitness \bar{W} that illuminates multilocus dynamics by providing an intuitive, geometrical representation and an analytic tool. Letting $Z = W$ in (164) and reexamining the error terms in the proof of (164), we find

$$\Delta \bar{W} = \bar{W}^{-1} V_g + \bar{\Delta W} + 2 \sum_i u_i \Delta d_i + O(s^3), \quad t \geq t_1, \quad (169a)$$

as $s \rightarrow 0$, where (93) gives the genic variance in fitness as

$$V_g = 2 \sum_k \sum_{i_k} p_{i_k}^{(k)} \xi_{i_k}^{(k)} w_{i_k}^{(k)}, \quad (169b)$$

$$\bar{\Delta W} = \sum_{i,j} (\Delta W_{ij}) p'_i p'_j \quad (169c)$$

represents the mean of the changes in the genotypic fitnesses over the next generation, and u_i designates the epistatic deviation in (104).

The paragraph below (164) applies with the single obvious change that $u_i = 0$ for every i if fitnesses are additive between loci.

By (116), we can approximate the genic variance in fitness on the linkage-equilibrium surface:

$$V_g(\mathbf{p}, t) = V_g(\boldsymbol{\pi}, t) + O(s^3), \quad t_1 \leq t \leq K/s. \quad (170)$$

Therefore, we could replace $V_g(\mathbf{p}, t)$ in (169a) by its linkage-equilibrium approximation $V_g(\boldsymbol{\pi}, t)$ without changing the order of the error.

Under the assumption (68) of weak explicit time dependence, (70) holds, and consequently (169a) simplifies to

$$\Delta \bar{W} = \bar{W}^{-1} V_g + \bar{\Delta W} + O(s^3), \quad t \geq t_2. \quad (171)$$

Under the further assumption that the genotypic fitnesses vary sufficiently slowly to satisfy

$$\Delta W_{ij} = O(s^3) \quad (172)$$

for every i and j , we obtain the asymptotic fundamental theorem of natural selection:

$$\Delta \bar{W} = \bar{W}^{-1} V_g + O(s^3), \quad t \geq t_2. \quad (173)$$

In particular, (173) holds for constant genotypic fitnesses. The weaker assumption $\Delta W_{ij} = o(s^2)$ leads to (173) with $o(s^2)$ instead of $O(s^3)$.

The results (169) (with D_i instead of d_i) to (173) were established for two multiallelic loci in NAGYLAKI (1976, 1977b, Sec. 8.2; 1992a, Sec. 8.2).

Thus, after the short time t_2 defined in (70), the *absolute* error in (173) is only $O(s^3)$. Since $V_g = O(s^2)$, we expect generically

$$\Delta \bar{W} \approx \bar{W}^{-1} V_g > 0, \quad t \geq t_2, \quad (174)$$

which would imply increase of the mean fitness. However, (174) can fail if the *relative* error in (173) is not small, which can occur if $V_g(\mathbf{p}, t) \ll s^2$. But (169b) and (15) yield

$$V_g(\boldsymbol{\pi}, t) = 2 [\bar{W}(\boldsymbol{\pi}, t)]^2 \sum_k \sum_{i_k} (\Delta \pi_{i_k}^{(k)})^2 / \pi_{i_k}^{(k)} \quad (175)$$

on the linkage-equilibrium surface, which shows that $V_g(\boldsymbol{\pi}, t) \ll s^2$ if and only if $|\Delta \pi_{i_k}^{(k)}| \ll s$ for every k and i_k . From (53) and (170) we conclude that $V_g(\mathbf{p}, t) \ll s^2$ if and only if $|\Delta p_{i_k}^{(k)}| \ll s$ for every k and i_k . Such slow gene-frequency change can result from symmetry conditions or proximity to an equilibrium (MORAN 1964; KIMURA 1965; NAGYLAKI 1977a).

If the fitnesses are constant and the gametic frequencies converge to an equilibrium point, the total changes in \bar{W} during the various evolutionary epochs are given by the estimates in NAGYLAKI (1977a).

We shall present two short proofs of (164). The first leads directly to the genic covariance C_g . If the population is initially in linkage equilibrium, however, the response to the first generation of selection involves the gametic variance C_{gam} (NAGYLAKI 1992b). The second proof is based on (119) and demonstrates that, to the leading order in s , the contributions of the epistatic variance and the effect of linkage disequilibrium add up to the sum in (164a). This proof generalizes the one for two multiallelic loci (NAGYLAKI 1976, 1977b, Sec. 8.2; 1992a, Sec. 8.2).

First proof of (164): From (85c), (86a), and (86b) we easily find

$$\Delta \bar{Z} = \bar{\Delta Z} + 2 \sum_i z_i \Delta p_i + \sum_{i,j} z_{ij} (\Delta p_i) \Delta p_j, \quad (176)$$

where $\bar{\Delta Z}$ is given by (164b). By (40), for $t \geq t_1$ the second sum is $O(s^2)$. To calculate the first sum, in successive lines we use (87), (18a); (4), (18b), and (24); (15) and (18a); and (99a), (89), (22), and (39):

$$\begin{aligned} 2 \sum_i z_i \Delta p_i &= 2 \sum_i \left(\sum_k \zeta_{ik}^{(k)} + e_i \right) \Delta p_i \\ &= 2 \sum_k \sum_{ik} \zeta_{ik}^{(k)} \sum_i \Delta p_i + 2 \sum_i e_i (\Delta d_i + \Delta q_i) \\ &= 2 \sum_k \sum_{ik} \zeta_{ik}^{(k)} \Delta p_{ik}^{(k)} + 2 \sum_i e_i \Delta d_i \\ &\quad + 2 \sum_i e_i \sum_k (q_i / p_{ik}^{(k)}) \Delta p_{ik}^{(k)} + O(s^2) \\ &= 2 \bar{W}^{-1} \sum_k \sum_{ik} p_{ik}^{(k)} \zeta_{ik}^{(k)} w_{ik}^{(k)} + 2 \sum_i e_i \Delta d_i \\ &\quad + 2 \bar{W}^{-1} \sum_k \sum_{ik} w_{ik}^{(k)} \sum_i \zeta_{ik}^{(k)} (p_i - d_i) e_i + O(s^2) \\ &= \bar{W}^{-1} C_g + 2 \sum_i e_i \Delta d_i + O(s^2), \quad t \geq t_1. \quad (177) \end{aligned}$$

Second proof of (164): We proceed as in the above proof, but evaluate the first sum in (176) differently. In successive lines, we employ (11a); (105a) and (87); (106); and (119) and (112):

$$\begin{aligned} 2 \sum_i z_i \Delta p_i &= 2 \sum_i z_i (\bar{W}^{-1} p_i w_i - D_i) \\ &= \bar{W}^{-1} C_{\text{gam}} - 2 \sum_i \left(\sum_k \zeta_{ik}^{(k)} + e_i \right) D_i \\ &= \bar{W}^{-1} (C_g + C_e) - 2 \sum_i e_i D_i \\ &\quad - 2 \sum_k \sum_{ik} \zeta_{ik}^{(k)} \sum_i D_i \\ &= \bar{W}^{-1} C_g + 2 \sum_i e_i \Delta d_i + O(s^2), \quad t \geq t_1. \quad (178) \end{aligned}$$

DISCUSSION

Here, we reference and summarize our main results and discuss some open problems. In the abstract, we

outlined the evolution of the multilocus system under weak selection. At the beginning of each section, we presented and discussed our major results in detail before proving them in the following subsections. Our results hold if $s \ll c_{\min}$, where s and c_{\min} denote the selection intensity, defined by (20), and the smallest two-locus recombination frequency, respectively. The numerical studies of LEWONTIN (1964a,b) and FRANKLIN and LEWONTIN (1970) do not satisfy this condition.

The bound (37) and the estimate (39) describe the rapid reduction of the linkage disequilibria d_i , defined by (18), to $O(s)$ in an evolutionarily short time $t_1 \sim (\ln s) / \ln(1 - c_{\min})$. According to (41), the estimate (39) holds also for the linkage disequilibria D_i , defined by (11b). The linkage disequilibria $\Delta_{iR}^{(R)}$ are defined by (128) for all subsets of loci R , and the explicit formula (142) shows that they, too, satisfy (39). The shortness of the time t_1 implies that whenever $s \ll c_{\min}$, linkage disequilibria in natural populations should almost always be small. This conclusion agrees with observation (HEDRICK 1983, pp. 386–393; KIMURA 1983, pp. 284–285; NEI 1987, pp. 171–175).

Equation 53 demonstrates that after the time t_1 , the population evolves approximately as if it were in linkage equilibrium, the error in the gametic frequencies being $O(s)$.

Suppose now that the explicit time dependence (if any) of the genotypic fitnesses is $O(s^2)$, as specified by (68). Then the single-generation change $\Delta d_{iR}^{(R)}$ in the linkage disequilibrium $d_{iR}^{(R)}$ satisfies the bound (69) for every subset R , and this yields the estimate (70): $\Delta d_i = O(s^2)$ for $t \geq t_2 \sim 2t_1$. As noted below (142), the linkage disequilibria $\Delta_{iR}^{(R)}$ also satisfy (70). As remarked below (148), provided the position effect (if any) is $O(s^2)$ [see (135)], the estimate (70) applies as well to the linkage disequilibria $D_{iR}^{(R)}$. Thus, all three sets of linkage disequilibria are nearly constant after the evolutionarily short time t_2 .

The gametic variance in an arbitrary character Z and the gametic covariance between Z and the fitness W were decomposed orthogonally into their additive and epistatic components, and the total epistatic variance and covariance, which involve only additive interactions, were further decomposed orthogonally into the contributions of all the subsets of loci. As indicated in (112) to (116), for $t \geq t_1$ these analyses can be approximated on the linkage-equilibrium surface.

According to (15), gene-frequency change is controlled by the average allelic excesses $w_{ik}^{(k)}$. Equations 118, 123, 124, and 126 yield

$$\Delta d_{iR}^{(R)} = \bar{W}^{-1} p_{iR}^{(R)} u_{iR}^{(R)} - D_{iR}^{(R)} + O(s^2), \quad t \geq t_1, \quad (179)$$

which reveals that for $t \geq t_1$, the evolution of the linkage disequilibria $d_{iR}^{(R)}$ is governed primarily by

$u_{i_R}^{(R)}$, the sum of the effects $v_{i_S}^{(S)}$ of the gametic additive fitness interactions in all subsets $S \subseteq R \subseteq N$ of loci. Thus, dominance deviations and their interactions do not affect the dynamics, and a diploid analysis of variance is unnecessary.

For $t \geq t_2$, (125) relates the linkage disequilibria $D_{i_R}^{(R)}$ to the epistatic deviations $u_{i_R}^{(R)}$. If the position effect is $O(s^2)$, then $D_{i_R}^{(R)}$ and $\Delta_{i_R}^{(R)}$ are given by (148) and (149), respectively, and (136) relates the linkage disequilibria $\Delta_{i_R}^{(R)}$ to $v_{i_R}^{(R)}$, the effect of the gametic additive fitness interaction of all the loci in R . Therefore, whenever the gametic frequencies and the linkage map are known, (136) can be used to bound or even estimate the epistatic deviations in fitness. Since the time t_2 is usually short, such a bound or estimate does not depend on any assumption of stationarity.

Our most general approximations for the rates of change $\Delta \bar{Z}$ and $\Delta \bar{W}$ of the means of \bar{Z} and \bar{W} are (164) and (169), respectively. Under the mild restriction (68) on the explicit time dependence of the genotypic fitnesses W_{ij} , these approximations simplify to (166) and (171), respectively. If, in addition, the genotypic trait values Z_{ij} vary sufficiently slowly to satisfy (167), we obtain the asymptotic secondary theorem of natural selection in the simple form (168). For the mean fitness, the corresponding assumption (172) leads to the precise statement (173) of the asymptotic fundamental theorem of natural selection.

Next, we turn to some open problems.

The conditions under which (174) fails because the relative error in (173) is large are still insufficiently characterized even for two loci.

In this paper, we fixed the number of loci (n), and consequently did not determine the dependence of our bounds and estimates on n . This observation leads to several difficult questions that are particularly important in quantitative genetics, where we often wish to study the limit $n \rightarrow \infty$. Taking this limit in our results requires uniformity in n . With the extremely conservative definition (20) of the selection intensity s , uniformity seems plausible, but it may be very hard to prove. Furthermore, it may sometimes be unlikely that s remains small when n is very large, as exemplified by a phenotypic fitness function that is very small for extreme phenotypes.

These possible limitations would be removed if the condition $s \ll c_{\min}$ could be weakened to $s^* \ll c_{\min}$, where s^* designates the largest single-locus selection coefficient. To define the constant s^* precisely, we replace the genotypic fitnesses $W_{ij}(\mathbf{p}, t)$ in (20a) by the single-locus marginal fitnesses $W_{i/jk}^{(k)}(\mathbf{p}, t)$ and set

$$r_k(\mathbf{p}, t) = \left[\max_{i_k/jk} W_{i_k/jk}^{(k)}(\mathbf{p}, t) - \min_{i_k/jk} W_{i_k/jk}^{(k)}(\mathbf{p}, t) \right] / \min_{i_k/jk} W_{i_k/jk}^{(k)}(\mathbf{p}, t), \quad (180a)$$

$$s^* = \sup_{k, \mathbf{p}, t} r_k(\mathbf{p}, t). \quad (180b)$$

TURELLI and BARTON (1990) investigated weak selection on a character determined purely additively by many loci. The above considerations would be essential for a rigorous derivation and generalization of their results from our approach.

I thank WARREN EWENS for helpful comments on the manuscript and MITZI NAKATSUKA for typing it conscientiously and beautifully. This work was supported by National Science Foundation grant BSR-9006285.

LITERATURE CITED

- AKIN, E., 1979 *The Geometry of Population Genetics*. Springer, Berlin.
- AKIN, E., 1982 Cycling in simple genetic systems. *J. Math. Biol.* **13**: 305–324.
- AKIN, E., 1983 Hopf bifurcation in the two-locus genetic model. *Mem. Am. Math. Soc.* **44**: No. 284.
- ATKINSON, F. V., G. A. WATTERSON, and P. A. P. MORAN, 1960 A matrix inequality. *Q. J. Math. Oxford Ser.* **11**: 137–140.
- BARTON, N. H., 1986 The effects of linkage and density-dependent regulation on gene flow. *Heredity* **57**: 415–426.
- BARTON, N. H., and M. TURELLI, 1991 Natural and sexual selection on many loci. *Genetics* **127**: 229–255.
- BENNETT, J. H., 1954 On the theory of random mating. *Ann. Eugen.* **18**: 311–317.
- BODMER, W. F., 1965 Differential fertility in population genetics models. *Genetics* **51**: 411–424.
- BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- CROW, J. F., and T. NAGYLAKI, 1976 The rate of change of a character correlated with fitness. *Am. Nat.* **110**: 107–213, 400.
- EWENS, W. J., 1969a A generalized fundamental theorem of natural selection. *Genetics* **63**: 531–537.
- EWENS, W. J., 1969b Mean fitness increases when fitnesses are additive. *Nature* **221**: 1076.
- EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer, Berlin.
- EWENS, W. J., 1989 An interpretation and proof of the fundamental theorem of natural selection. *Theor. Popul. Biol.* **36**: 167–180.
- FELSENSTEIN, J., 1965 The effect of linkage on directional selection. *Genetics* **52**: 349–363.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- FLEMING, W. H., 1979 Equilibrium distributions of continuous polygenic traits. *SIAM J. Appl. Math.* **36**: 148–168.
- FRANKLIN, I., and R. C. LEWONTIN, 1970 Is the gene the unit of selection? *Genetics* **65**: 707–734.
- GEIRINGER, H., 1944 On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* **15**: 25–57.
- HASTINGS, A., 1981 Stable cycling in discrete-time genetic models. *Proc. Natl. Acad. Sci. USA* **78**: 7224–7225.
- HASTINGS, A., 1985 Multilocus population genetics with weak epistasis. I. Equilibrium properties of two-locus two-allele models. *Genetics* **109**: 799–812.
- HASTINGS, H., 1986 Multilocus population genetics with weak epistasis. Equilibrium properties of multilocus models: what is the unit of selection? *Genetics* **112**: 157–171.
- HASTINGS, A., 1989 Deterministic multilocus population genetics: an overview, pp. 27–54 in *Some Mathematical Questions in Biology: Models in Population Biology*, edited by A. HASTINGS. American Mathematical Society, Providence, R.I.
- HEDRICK, P. W., 1983 *Genetics of Populations*. Science Books International, Boston.

- KARLIN, S., 1975 General two-locus selection models: some objectives, results and interpretations. *Theor. Popul. Biol.* **7**: 364–398.
- KARLIN, S., 1978 Theoretical aspects of multi-locus selection balance. I, pp. 503–587 in *Studies in Mathematics, Vol. 16: Studies in Mathematical Biology*, edited by S. A. LEVIN. Mathematical Association of America, Washington, D.C.
- KARLIN, S., and M. W. FELDMAN, 1970 Convergence to equilibrium of the two locus additive viability model. *J. Appl. Prob.* **7**: 262–271.
- KIMURA, M., 1965 The attainment of quasi-linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* **52**: 875–890.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- KINGMAN, J. F. C., 1961 On an inequality in partial averages. *Q. J. Math. Oxford Ser.* **12**: 78–80.
- KUN, L. A., and YU. I. LYUBICH, 1979 Convergence to equilibrium under the action of additive selection in a multilocus multiallelic population. *Soviet Math. Dokl.* **20**: 1380–1382.
- LANDE, R., 1976 Natural selection and genetic drift in phenotypic evolution. *Evolution* **30**: 314–334.
- LANGLEY, C. H., and J. F. CROW, 1974 The direction of linkage disequilibrium. *Genetics* **78**: 937–941.
- LEWONTIN, R. C., 1964a The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.
- LEWONTIN, R. C., 1964b The interaction of selection and linkage. II. Optimum models. *Genetics* **50**: 757–782.
- LYUBICH, YU. I., 1971 Basic concepts and theorems of the evolutionary genetics of free populations. *Russ. Math. Surv.* **26**: 51–123.
- MOODY, M., 1978 A multi-locus continuous-time selection model. *J. Math. Biol.* **5**: 281–291.
- MORAN, P. A. P., 1964 On the nonexistence of adaptive topographies. *Ann. Hum. Genet.* **27**: 383–393.
- MORAN, P. A. P., 1967 Unsolved problems in evolutionary theory. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.* **4**: 457–480.
- MULHOLLAND, M. P., and C. A. B. SMITH, 1959 An inequality arising in genetical theory. *Am. Math. Month.* **66**: 673–683.
- NAGYLAKI, T., 1976 The evolution of one- and two-locus systems. *Genetics* **83**: 583–600.
- NAGYLAKI, T., 1977a The evolution of one- and two-locus systems. II. *Genetics* **85**: 347–354.
- NAGYLAKI, T., 1977b *Selection in One- and Two-Locus Systems*. Springer, Berlin.
- NAGYLAKI, T., 1979a Selection in dioecious populations. *Ann. Hum. Genet.* **43**: 143–150.
- NAGYLAKI, T., 1979b Dynamics of density- and frequency-dependent selection. *Proc. Natl. Acad. Sci. USA* **76**: 438–441.
- NAGYLAKI, T., 1987 Evolution under fertility and viability selection. *Genetics* **115**: 367–375.
- NAGYLAKI, T., 1989 Rate of evolution of a character without epistasis. *Proc. Natl. Acad. Sci. USA* **86**: 1910–1913.
- NAGYLAKI, T., 1991 Error bounds for the fundamental and secondary theorems of natural selection. *Proc. Natl. Acad. Sci. USA* **88**: 2402–2406.
- NAGYLAKI, T., 1992a *Introduction to Theoretical Population Genetics*. Springer, Berlin.
- NAGYLAKI, T., 1992b Rate of evolution of a quantitative character. *Proc. Natl. Acad. Sci. USA* **89**: 8121–8124.
- NAGYLAKI, T., and J. F. CROW, 1974 Continuous selective models. *Theor. Popul. Biol.* **5**: 257–283.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- PASSEKOV, V. P., 1984 Asymptotic analysis of selection in a multilocus, multiallelic population (in Russian). *Dokl. Akad. Nauk SSSR* **277**: 1338–1341.
- PRICE, G. R., 1972 Fisher's "fundamental theorem" made clear. *Ann. Hum. Genet.* **36**: 129–140.
- REIERSÖL, O., 1962 Genetic algebras studied recursively and by means of differential operators. *Math. Scand.* **10**: 25–44.
- ROBERTSON, A., 1966 A mathematical model of the culling process in dairy cattle. *Anim. Prod.* **8**: 95–108.
- ROBERTSON, A., 1968 The spectrum of genetic variation, pp. 5–16 in *Population Biology and Evolution*, edited by R. C. LEWONTIN. Syracuse University Press, Syracuse.
- SCHEUER, P. A. G., and S. P. H. MANDEL, 1959 An inequality in population genetics. *Heredity* **13**: 519–524.
- SHAHSHAHANI, S., 1979 A new mathematical framework for the study of linkage and selection. *Mem. Am. Math. Soc.* **17**: No. 211.
- SLATKIN, M., 1972 On treating the chromosome as the unit of selection. *Genetics* **72**: 157–168.
- TURELLI, M., and N. H. BARTON, 1990 Dynamics of polygenic characters under selection. *Theor. Popul. Biol.* **38**: 1–57.

Communicating editor: R. R. HUDSON